Analysis and Applications of Complex Social Networks 2018

Lead Guest Editor: Katarzyna Musial
Guest Editors: Piotr Brodka and Pasquale De Meo

Analysis and Applications of Complex Social Networks 2018

Analysis and Applications of Complex Social Networks 2018

Lead Guest Editor: Katarzyna Musial
Guest Editors: Piotr Brodka and Pasquale De Meo

Editorial Board

Contents

Editorial

Analysis and Applications of Complex Social Networks 2018

Katarzyna Musial [iD],[1] **Piotr Bródka** [iD],[2] **and Pasquale De Meo** [iD][3]

[1]*University of Technology Sydney, Sydney, Australia*
[2]*Wroclaw University of Science and Technology, Wroclaw, Poland*
[3]*University of Messina, Messina, Italy*

Correspondence should be addressed to Katarzyna Musial; katarzyna.musial-gabrys@uts.edu.au

It is our great pleasure to present to you the second edition of this special issue discussing the analysis and applications of complex social networks. Similarly to the one published in this journal last year, this one also turned out to be a great success as we managed to attract a number of high-quality researches in the area of complex social networks.

The research space in complex social networks grows every year as they are systems with many levels of complexity and there is a constant need to challenge our current understanding in the field. The results of the community research efforts enable the understanding of different social phenomena including social structures evolution, communities, spread over networks, and control in and of complex networks. This huge interest in the analysis of large-scale social networks resulted in a lot of new approaches, methods, and techniques but with every advancement in this area, we uncover new challenges and new levels of complexity in the network universe that are far from being explored and addressed. The increasing complexity of the tasks to be performed in terms of network analysis together with the volume, variety of social data about people and their interactions, and velocity with which this data is generated in the online world poses new requirements and challenges on researchers. One of them is how to build accurate methods that would be able to cope with these vast amounts of data. This issue is a result of an attempt to address these emerging challenges with a big emphasis on the applicability of the developed approaches.

One of the goals of this special issue is to show that analysis of large-scale, real-world social networks underpinned by

fundamental research is the direction to take when it comes to the future of complex social network analysis. We emphasize that in the world of network science fundamental research and application- and data-driven research are equally important and they need to go together to generate significant academic, societal, and commercial impact.

The variety of papers published in this special issue shows that there is a long list of topics that have not yet been comprehensively researched. These papers also show the future challenges and trends in analysis and applications of complex social networks. Within this special issue, we present a wide variety of application-driven studies looking, for example, at complexity of a microblogging system, transportation systems, an emergency management system, organizational structure and management, innovation, or food safety. The fundamental research that is covered within these special issues ranges between (i) investigation and analysis of network structure and metrics, e.g., signed networks, modularity, and communities, (ii) link prediction approaches, (iii) resilience in complex networks, (iv) diffusion and influence, and (v) control in networked systems—the topic that is currently of great interest to the community.

Some of the papers already in this special issue are as follows: "The Settlement Structure Is Reflected in Personal Investments: Distance-Dependent Network Modularity-Based Measurement of Regional Attractiveness" by L. Gadar et al.; "Simulation of Knowledge Transfer Process Model Between Universities: A Perspective of Cluster Innovation Network" by F. Wei and X. Limin; "Variational Approach for Learning Community Structures" by J. J. Choong et

al.; "Complexity of a Microblogging Social Network in the Framework of Modern Nonlinear Science" by A. Dmitriev et al.; "More on Spectral Analysis of Signed Networks" by G. Yu and H. Qu; "Modelling Multilevel Interdependencies for Resilience in Complex Organisation" by J. Tasic et al.; "Establishment and Analysis of the Supernetwork Model for Nanjing Metro Transportation System" by Y. Wei and S. Ning.

This special issue also contains the following papers: "A Semantic Community Detection Algorithm Based on Quantizing Progress" by X. Han et al.; "Scare Behavior Diffusion Model of Health Food Safety Based on Complex Network" by J. Luo et al.; "Examining the Intergovernmental and Interorganizational Network of Responding to Major Accidents for Improving the Emergency Management System in China" by P. Tang et al.; "Exponential Synchronization Control of Discontinuous Nonautonomous Networks and Autonomous Coupled Networks" by C. Yang et al.; "Crisis Spreading Model of the Shareholding Networks of Listed Companies and Their Main Holders and Their Controllability" by Y. Ma and L. Li; "Predicting Missing Links Based on a New Triangle Structure" by S. Bai et al.; "Competition-Based Benchmarking of Influence Ranking Methods in Social Networks" by A. Topîrceanu.

Published papers show that although all of the presented topics have been researched for many years now, there is still space and need for new contributions. Challenges change their nature as we face vast amounts of heterogeneous data that are continuously generated. Network resilience, communities, spread and influence analysis, network complexity, control, and structural properties are topics that are trending in the research community all over the world. Those are very hard problems to address because of their complexity originating from two sources: (i) system: variety of connections, attributes of nodes and connections, nontrivial structure, and dynamics of a system; (ii) process: evolution driven by a variety of factors including external ones that are very hard to capture, spreading over complex structure of multiple processes or needed process adaptation connected with evolving structure. Thus, there is a continuous need to create cross-disciplinary teams that would work on those challenges with a holistic view of the problem.

So our work does not stop here, and we aim at continuing to bring together people from different fields to work on the topics covered within this special issue.

Conflicts of Interest

The editors declare that they have no conflicts of interest regarding the publication of this special issue.

Acknowledgments

This special issue is an outcome of the hard work of a number of people and could not have happened without the support of our collaborators. We would like to thank the editors-in-chief of this journal for their kind support and help during the entire process of publication. This was possible thanks to the work of the researchers who provided their anonymous reviews. Finally, we are most grateful to the authors for their valuable contributions and for their willingness and efforts to improve their papers in accordance with the reviewers' suggestions and comments.

Katarzyna Musial
Piotr Bródka
Pasquale De Meo

Research Article

Modelling Multilevel Interdependencies for Resilience in Complex Organisation

Justyna Tasic,[1,2,3] **Fredy Tantri,**[1,2,3] **and Sulfikar Amir** [ID][3,4]

[1]*Interdisciplinary Graduate School, Nanyang Technological University, Singapore*
[2]*Institute of Catastrophe Risk Management, Nanyang Technological University, Singapore*
[3]*Future Resilient Systems, Singapore-ETH Centre, Singapore*
[4]*School of Social Sciences, Nanyang Technological University, Singapore*

Correspondence should be addressed to Sulfikar Amir; sulfikar@ntu.edu.sg

This paper aims to model multilevel interdependencies in complex organisational systems and proposes application for resilience analysis. Most of the existing research studied interdependencies only at the single-level and overlooked their multilevel character. In response to this gap, we propose a multilevel approach to better comprehend the complexity of interdependencies in organisational systems. More specifically, the study focuses on how interdependencies are shaped across multiple organisational levels. To understand the research problem, we use multilevel and social network theories to elaborate the concept at five organisational levels, namely, individual, intraunit, interunit, intraorganisational, and interorganisational. Further, we show the application of multilevel interdependencies into analysis of organisational resilience. To this end, we construct a multiplex model of a real world organisational system that comprises formal and informal relations of different social exchange strength. Using the agent-based simulations of the organisational system, we investigate the relations between organisational interdependencies and organisational performance in normal and disrupted conditions. With the results, we argue that managing multilevel interdependencies is crucial to reduce vulnerability of organisational systems. By introducing the multilevel conceptualisation of interdependencies and presenting their influence on organisational resilience, we hope to pave a path to managing the complexity of interdependencies and strategic resilience enhancement in organisational systems.

1. Introduction

What are the consequences of interdependencies that occur at multiple levels on our capacity for responding to disruptions? This paper examines the emergence of multilevel interdependencies and how they affect organisational behaviour in mitigating crisis. Organisational behaviours have their origins and consequences at different levels [1, 2]. In analysing this correlation, most management researchers study individuals, work groups, organisations, and industry using approaches in which the structure of interactions of these entities is typically observed only at one level [3]. Research conducted at the micro-, meso-, and macro-levels often ignores the fact that organisational dynamics result from multilevel interactions. This particular aspect is crucial in understanding the impact of interdependencies on organisational performance.

In management science, interdependencies are understood as exchange relationships [4, 5] within and between organisations [6]. In general, studies on interdependencies are divided between those looking at the micro- and mesolevel and those at the macro one. Studies on the micro- and mesointerdependencies are firmly grounded in organisational behaviour and social science theories, which have been used to explain individual or group behaviour in the context of task interdependence, workflows, and interpersonal or intergroup relations [7–11]. The macro approach has been usually adopted by economics, sociology, strategy, and supply chain management scholars. This approach provides the understandings of interorganisational relations between companies, supply chain partners, and regulatory and institutional entities that affect organisational performance [6, 12–17]. While single-level approaches have their own virtues,

there is a gap of knowledge in how these levels interact with one another and what implications they bring on organisational resilience. Except for the divergent focuses, research on interdependencies at single levels often uses structural-functionalism and social exchange theory lenses [14, 18–24] to investigate interdependencies. However, due to disparate goals, organisational experts of each of the levels have seldom merged their perspectives to achieve a bigger picture of organisational dynamics. As a result, adopting either micro-, meso-, or macro-lens only yields an incomplete understanding of multilevel interactions and their consequences. It is this lack of knowledge in multilevel organisational interdependencies that this paper seeks to shed light on.

Grounded in general systems theory [25, 26], a multilevel view on organisational analysis differs from analysing organisations through a single-level lens. It aims to shift the research field toward the view of organisations as complex and interconnected social systems [27, 28] where the phenomena are analysed at interlevel domains. Because interdependencies in organisational systems are too intricate phenomenon to be explained by only single-level terms, their comprehensive understanding requires integration of micro-, meso-, and macro-insights. Nowadays, due to economic and technological changes, the degree of interdependencies within and between organisations constantly grows [29, 30]. In this light, using the structural approach, we aim to introduce a preliminary conceptualisation of multilevel interdependencies in organisational systems and present its applications for resilience analysis. We elaborate the concept using the *multilevel* [31] and *social network theories* [32–34], which help us to explain in detail how interdependencies are formed and how they can be measured at each level. We argue that presenting interdependencies as a multilevel construct enables a more integrated and comprehensive understanding of the phenomenon's complexity and dynamics that unfold across organisational levels, and affect organisational resilience. We prove it with agent-based simulations of a real organisational system under the crisis condition.

To achieve this objective, this paper is structured as follows. The following two sections present the state-of-art in interdependencies and resilience studies of organisational systems, in which we highlight the key aspects of interdependencies and resilience studies and identify areas where multilevel analysis remains lacking. After reviewing the existing works on organisational interdependencies and resilience, we begin to introduce a multilevel conceptualisation of interdependencies in complex organisational systems. We move then to the empirical application of the concept in the area of resilience analysis. More specifically, we construct a multiplex model of a real world organisational system that comprises weighted formal and informal relations between and within two organisations. By giving the weights to multiplex relations we aim to reflect their different social exchange strength. We establish the weights based on the system of informal relationships introduced by Luo [35] and the hierarchical logic of formal employment relationships, which determine chains of command (authority) and formal information exchanges (reporting) [36]. We also describe in details qualitative and quantitative characteristics of existing

edges. Further, we present the relations between multilevel interdependencies and organisational performance in normal and disrupted conditions. Next, we discuss the results and highlight that managing multilevel interdependencies is crucial to reduce vulnerability and enhances resilience of organisational systems. The conclusion summarises theoretical and empirical contributions of the paper.

2. Linking Interdependencies to Resilience

Interdependence began to be a subject of inquiry for organisational research in 1949 when Deutsch introduced a theory of social interdependence, which gained influences on management research and practices for several decades. Since then, organisation researchers have examined the nature and consequence of interdependence from various perspectives. A dichotomy between micro and macro levels appears to characterize the study of social interdependencies. In the micro studies of interdependencies, La Porte [4] defined the degree of organised system complexity as a function of number of components, differentiation of components, and interdependent links between the components. In similar direction, management science explored interdependencies with the theory firmly grounded in organisational psychology and social science to explain human behaviour in the context of task interdependence, interpersonal relationships and performance at micro- and mesolevels of analysis [7–11]. At the macro level, analysis of interdependencies in management studies drew theoretical underpinnings on disciplines such as economics, economic sociology, and supply chain management to assess interorganisational relations that shape organisational performance, for instance, flows of goods and services, trust-based relationships, and industry environment [6, 12–16]. In this area of research, only a few attempts looked into interdependencies across levels [37, 38].

In further developments of organisational sciences, interdependencies are understood as social exchange relationships [4, 5] within and between organisations [6]. At micro level, Brass [21] underlined that an organisation is composed of interdependent networks of employees. At mesolevel, the interdependencies are defined as a relationship between two or more organisational units that are mutually affected [19]. Following this path, Gulati [39] and Tomkins [14] tracked down the source of interdependencies in the interorganisational context by focusing on the role of trust and information relations as determinants of the interdependent relationship. Conceptually, the notion of interdependence assumes at least one exchanged resource per interdependent dyad [4]. However, in real life, the relations become more complex, due to the multiplex types of independencies. It is extremely important to note that while organisational relations are *formally* maintained, through the existence of rules and regulations [6, 40, 41], they are also *informally* shaped, through social structures, trust, and reciprocity [23, 40, 42]. Informal relations embody interpersonal trust, which lubricates social support and, thus, enables complex transactions and facilitates collective action. Therefore, structural position of actor in an organisation is formed by both formal and emergent informal interdependencies, which interact with

each other. In this area, interdependencies were analysed as networks of formal work exchanges [e.g., [21, 43]], and informal relationships as acquaintance, friendship and familiarity networks [35, 44, 45]. While the formal relationships are easily identified, the informal interactions are hardly detectable and often overlooked; however, they do impact organisational performance [45–48].

There are two commonly distinguished types of ties—instrumental and expressive [49]. Instrumental ties are information and cognition based exchanges of resources for instrumental purposes (e.g., reporting, acquaintance). Expressive ties include an affective factor and reflect common relationship identity as well as social support (e.g., friendship). In addition, researchers also recognised existence of mixed ties, which combine features of both types [35]. According to Luo [35], informal instrumental ties are based on rules of equity (acquaintance), informal expressive ties are governed by rules of need (friendship), and mixed ties by rules of favour (familiarity). Both strong instrumental and expressive ties have a positive effect on performance; in contrast weak instrumental ties are less likely to enhance performance [50].

Overall, the research on organisational interdependencies has produced a growing body of knowledge to illuminate the dynamics of organisational systems as a result of complex interactions and ever expanding structures. However, as our review of this body of knowledge above indicates, most of the research is inclined to pay attention only at interactions and processes that occur at the same level. Characters and properties of interdependencies that emerge out of multilevel dynamics remain to be explored. In addition, the influence of informal interactions on organisational functioning requires in-depth analysis.

It is posited in this paper that the multilevel structure of interdependencies has potential impact on organisational capacity for responding to crisis, especially large-scale ones. What remains unclear is how exactly structural interdependencies will affect organisational resilience. To have a clear explanation, we first need to have a basic understanding of what constitutes resilience and what renders organisation to be resilient. Up to date, the concept of resilience has been applied in a wide variety of academic fields, including among others, ecology [51–53], sociology [54], psychology [55], supply chain management [56], strategic management [57], disaster mitigation [58, 59], sociotechnical resilience [60, 61], resilience engineering [62, 63], organisational reliability [29, 64, 65], and resilience management [66–68]. In the following lines, we describe the concept of resilience with a primary reference to resilience engineering and management studies, which constitute the most relevant ground for this paper.

Many scholars acknowledge that resilience offers a potential solution to overcome disruptions and uncertainties in organisational systems, and create the environment for organisational development. In 2006, a group of experts in industrial and system safety initiated an organisation-centred paradigm called "resilience engineering" [62]. Recently, the continuations of those developments led to the following definition of resilience: "Resilience is an expression of how people, alone or together, cope with everyday situations –

large and small – by adjusting their performance to the conditions [69]. Resilience in the field of management is generally regarded as "an emergent property that relates to the inherent and adaptive qualities that enable an organisation to take a proactive approach to threat and risk mitigation" [70]. In general terms, an organisational system "can be seen as more resilient when it is more robust and less vulnerable to disruptions and recovers faster from disruptions when they occur" [71]. To sum up, resilient organisational systems are proactive in mobilizing their resources, use their abilities and adapt to effectively perform under a variety of conditions (both expected and unexpected).

How do organisational structures affect resilience? As pointed out by a number of researchers, complex systems organisations are composed of the multiple levels of operation [72] in which interconnected agents form a network of nonlinear relationships that give rise to emergent behaviours [66]. The more complex organisation, the more complex response to disruptions will become [72]. Similar to safety, resilience is seen as a systemic property where individual, group, and organisational levels have a reciprocal influence on each other [73, 74]. From this vantage point, managing resilience is "as a matter of a balancing between individual resilience (individual responses to operational challenges) and system resilience (the large scale autocatalytic combination of individual behaviour)" [62]. Resilient employees and units do not guarantee resilient organisation. In addition, organisations are embedded in interorganisational relationship, which are crucial for sustaining their operations [13, 56, 75]. Organisations are more likely to be resilient if they can effectively mobilise internal and external resources and capabilities [74]. Therefore, interorganisational resilience is usually described as adaptability, flexibility and redundancy or diversification [56, 75]. Applying systems thinking approach, Leveson [63] argued that system safety is an emergent property and thus must be controlled at the system level. Therefore, building resilience should be seen as a collaborative process, which requires continuous learning and adaptation from all actors, i.e., across individuals, groups, organisations [68, 76], and interorganisational networks.

The bottom line is that organisational resilience is strongly related to everyday work and organisational structure [60, 65, 74, 77]. Well-developed relationships (formal and informal) within and among organisations are fundamental for taking up a joint action during normal and crisis operations [29, 45, 47, 65]. The informal control instruments can substitute or complement formal control instruments in both normal [14, 23, 78–80] and crisis operations [47]. Thus, any resilient organisational system has a strong culture of awareness and is able to create improvised responsive networks to mitigate a crisis [47, 59, 65, 67].

3. Multilevel Interdependencies in Complex Organisation

Multilevel theory in organisational studies has its conceptual underpinnings in general systems theory [25, 26], which presents organisations as complex and interconnected social systems [27, 28, 72, 81]. Complexity of organisational systems
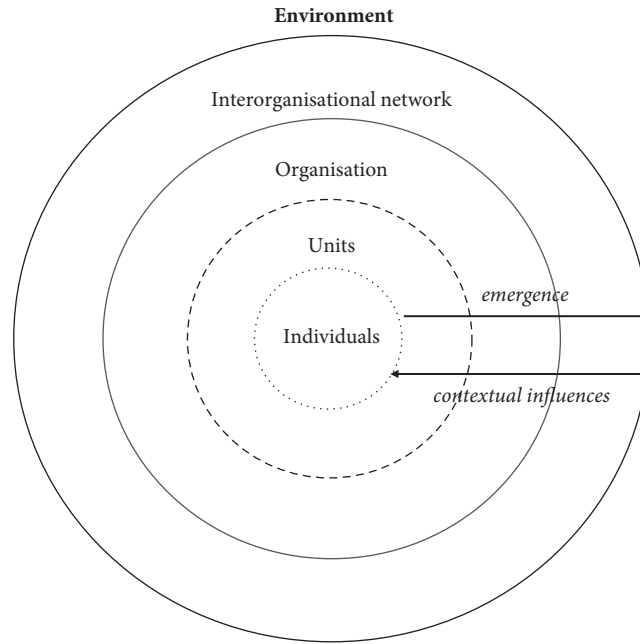
FIGURE 1: Nested character of complex organisational systems.

arises from interdependencies of the system components and emergent behaviour [82]. In this view, an organisation is conceptualized as a set of subsystems composed of more elemental components (Figure 1) that are arrayed in a hierarchical structure [25]. Organisational systems are seen as tightly coupled and nonlinear structures blended with bidirectional causal loops [83].

The main assumptions underlying the multilevel approach are that many phenomena emerge at multiple levels of analysis, and that organisational entities reside in nested arrangements [1–3]. The connection between levels of a system is determined by their couplings, which is the extent to which properties, dynamics, and processes at one level affect other levels [25]. Partial analysis of the system can be misleading, and thus to avoid wrong conclusions multilevel analysis is extremely crucial to represent all properties of the system [84]. The implication of the system approach is double-side. On one side, in line with general system theory, dynamics and relationships at lower levels emerge over time to higher levels of a system to yield structure [85]. This view coincides with theories of complexity, chaos, and self-organisation. On the other side, many organisational theories postulated that the contextual factors at higher levels may have direct or moderating effect on lower levels phenomena [31]. Although, we acknowledge that interdependencies, as many other phenomena in organisational systems, are a fuzzy concept that emerges bottom-up and is affected by top-down contextual processes, in this paper our focus is primarily on the emergent character of interdependencies in order to open a research discourse on multilevel character of interdependencies.

In advancing multilevel interdependencies informed by the theoretical insights described above, we use *configural compilation approach* to analyse interdependencies as a multilevel construct [1, 31, 83, 86]. The compilational approach

rests on the assumption of discontinuity and complex nonlinear processes of emergence. It describes phenomena that comprise a common domain but are distinctively different as they emerge across levels, i.e., simple aggregation of a construct at multiple levels is impossible [31]. In compilation processes, concept properties across all levels are discontinuous—qualitatively different— yet functionally equivalent—have the same role [1]. In this way, functional equivalence allows analysing a phenomenon by the elemental content of different types and amounts but possessing similar collective properties across levels [87]. Configural properties of a phenomenon are based on compilation models of emergence, i.e., phenomenon is nonlinear, not uniformly distributed, and not isomorphic across levels. They capture the differential patterns of relationship and variability of lower-level contributions to yield higher-level properties. This means unit-level interdependencies are a complex configuration of unique characteristics of unit members that emerge as a whole. In sum, with application of configural compilation approach, we argue that interdependencies across all levels are discontinuous phenomenon that does not express uniform pattern due to their specific nature.

To further elaborate this concept, the *social network theory* [32, 33, 89] is employed to explain in detail how relational exchanges comprising interdependencies are formed in organisations, and to depict their pattern at various levels. The pattern includes both formal and informal structure of relations that form organisational systems. Formal relations are prescribed set of interdependencies between employees established by the organisation that determine its formal functioning, i.e., authority and reporting relationships. Equally important to underline are informal networks that comprise of trust-based relationships such as acquaintance, friendship and familiarity networks. These informal
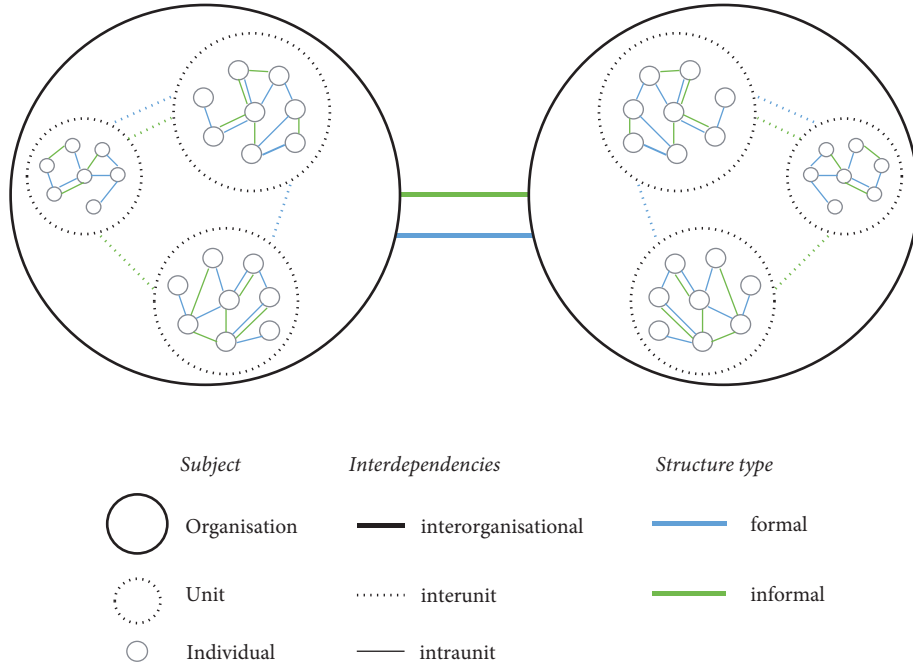
Figure 2: Overview of nested multilevel interdependencies.

relations influence on individuals, group, organisation and interorganisational network performance [11, 13, 15, 43, 48, 90, 91]. In addition, informal organisational structure very often supplements the formal relations, especially in crisis situations [45, 47].

Applying the system view above, we recognise interdependencies to be nested at five organisational levels: *individual, intraunit, interunit, intraorganisational* and *interorganisational.* This span of levels represents a hierarchical structure in which each level represents unique characteristics of interdependencies between individuals, units, and organisations. The interdependencies at each level of representation are constituted by formal and informal relations. The interdependencies are coupled across levels and their content is meaningfully related in the whole network of relations. While lower levels such as individual and unit are composed of more elemental components, higher-level interdependencies, especially organisational and interorganisational, are relatively inclusive and encompass characteristics of lower-levels. In doing so, we assume that interpersonal interdependencies constitute individual interdependencies that contribute to intra- and interunit level interdependencies; intraunit and interunit interdependencies contribute to intraorganisational-level interdependencies (Figure 2).

As elaborated above, each organisational system is a system of multiplex networks that comprises formal (reporting and authority) and informal (acquaintance, friendship, and familiarity) relations, which form interdependencies across all levels of analysis. In the following paragraphs, we use weighted degree centrality to decompose and describe the multilevel interdependencies in organisational systems. Determining the weights depends on the particular analysis needs and the organisations features such as industry and

organisational culture. In line with the configural compilation approach, the applied measures vary across the functional equivalents at multiple levels due to different characteristics of interdependencies across the levels (individual, intraunit, interunit, intraorganisational, and interorganisational network).

(1) *Individual interdependencies* entail formal and informal relational exchanges that constitute weighted degree of an individual, which is the focal point of analysis. Formal relations (F) comprise relationships predetermined by the organisation, i.e., authority and reporting relationships between the individual and other directly connected individuals. Formal weighted individual degree ($w_i^{[F]}$) is the sum of weights associated to all formal edges attached to individual i, where N is the total number of individuals in the interorganisational network m

$$w_i^{[F]} = \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}^{[F]} \tag{1}$$

Informal relations (I) capture emergent organisational features and are the by-product of formal daily interactions and interpersonal attachment. They consist of trust-based networks, which influence on the work behaviour of the individual, i.e., acquaintance, friendship, and familiarity networks [35, 44]. Informal individual degree ($w_i^{[I]}$) is the sum of weights associated to all informal edges attached to individual i

$$w_i^{[I]} = \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}^{[I]} \tag{2}$$

Formal and informal relations interact with each other and affect individual's behaviour. Therefore, the individual's structural position is the result of particular combination of both formal and informal interdependencies. The individuals bonded in the direct neighbourhood will be highly interdependent and analysis of interdependencies at this level helps to assess individuals' connectedness.

(2) *Intraunit interdependencies* emerge from the configuration of the unit members' interdependencies that give a comprehensive picture of interdependent relations within a unit. The formal (F) and informal (I) exchange relationships remain the same features as at the individual level; however, the interdependencies are more complex due to the increased number of individuals and relationships. Formal intraunit degree ($w_u^{[F]}$) of unit u is the sum of formal edges attached to individuals within unit u, where K_u is the number of individuals in unit u (3). Informal intraunit degree ($w_u^{[F]}$) of unit u is the sum of weights associated to informal edges attached to individuals within unit u (4).

$$w_u^{[F]} = \sum_{i=1}^{K_u} \sum_{\substack{j=1 \\ j \neq i}}^{K_u} w_{ij}^{[F]} \tag{3}$$

$$w_u^{[I]} = \sum_{i=1}^{K_u} \sum_{\substack{j=1 \\ j \neq i}}^{K_u} w_{ij}^{[I]} \tag{4}$$

To allow the comparability of the unit structures between the units, we construct also the average formal ($\langle w_u^{[F]} \rangle$) and informal ($\langle w_u^{[I]} \rangle$) unit degree of a unit u ((5) and (6)), which are calculated by normalization of intraunit degrees by the number of individuals in a unit u.

$$\left\langle w_u^{[F]} \right\rangle = \frac{w_u^{[F]}}{K_u} \tag{5}$$

$$\left\langle w_u^{[I]} \right\rangle = \frac{w_u^{[I]}}{K_u} \tag{6}$$

Analysis of interdependencies at this level helps to identify the crucial individuals on whom the unit performance is the most dependent both formally and informally. Formal structure is always well known by the individuals involved in the unit. The analysis of informal structure helps to identify hidden unit patterns and notice hidden needs and opportunities. Formal and informal relations should be well balanced to facilitate better the unit performance.

(3) *Interunit interdependencies* encompass formal and informal relations that emerge from the cross-unit relations between individuals. Formal interunit degree ($w_{u_{int}}^{[F]}$) of unit u is the sum of weights of formal edges between individuals in unit u and individuals in other units (K_s) in organisation o, where U_o is the number of units in organisation o (7). Informal interunit degree ($w_{u_{int}}^{[I]}$) of unit u is the sum of

weights of informal edges between individuals in unit u and individuals in other units (K_s) in organisation o (8).

$$w_{u_{int}}^{[F]} = \sum_{i=1}^{K_u} \sum_{\substack{s=1 \\ s \neq u}}^{U_o} \sum_{j=1}^{K_s} w_{ij}^{[F]} \tag{7}$$

$$w_{u_{int}}^{[I]} = \sum_{i=1}^{K_u} \sum_{\substack{s=1 \\ s \neq u}}^{U_o} \sum_{j=1}^{K_s} w_{ij}^{[I]} \tag{8}$$

Formal interunit relations determine how units operate and how much formally interdependent is their work. The units can be very independent (divisional structure), moderately interdependent (functional structure), or very interdependent (matrix). The informal relations, described in the social network theory as informal boundary spanning, emerge from informal interunit exchanges of information and support (acquaintance, friendship, and familiarity relations). Informal interunit relations crucial for a unit to raise unit effectiveness and gain access to external resources; however, they work the best when the unit is well connected externally and well as internally.

We construct also the average formal ($\langle w_{u_{int}}^{[F]} \rangle$) and informal ($\langle w_{u_{int}}^{[I]} \rangle$) interunit degrees of a unit u ((9) and (10)), which are calculated by normalization of interunit degrees by the number of individuals in a unit u.

$$\left\langle w_{u_{int}}^{[F]} \right\rangle = \frac{w_{u_{int}}^{[F]}}{K_u} \tag{9}$$

$$\left\langle w_{u_{int}}^{[I]} \right\rangle = \frac{w_{u_{int}}^{[I]}}{K_u} \tag{10}$$

(4) *Intraorganisational interdependencies* include formal and informal relationships within and between units of an organisation. Formal intraorganisational degree ($w_o^{[F]}$) of an organisation o is the sum of formal intra- and interunit degrees within this organisation (11). Informal intraorganisational degree ($w_o^{[I]}$) of an organisation o is the sum of informal intra- and interunit degrees within this organisation (12). The formal and informal intraorganisational degrees are crucial to assess internal connectedness of an organisation.

$$w_o^{[F]} = \sum_{u=1}^{U_o} w_{u_{int}}^{[F]} + \sum_{u=1}^{U_o} w_u^{[F]} \tag{11}$$

$$w_o^{[I]} = \sum_{u=1}^{U_o} w_{u_{int}}^{[I]} + \sum_{u=1}^{U_o} w_u^{[I]} \tag{12}$$

The average formal ($\langle w_o^{[F]} \rangle$) and informal ($\langle w_o^{[I]} \rangle$) intraorganisational degrees ((13) and (14)) are calculated by normalization of the degrees by the total number of individuals (L_o) in an organisation o.

$$\left\langle w_o^{[F]} \right\rangle = \frac{w_o^{[F]}}{L_o} \tag{13}$$

$$\left\langle w_o^{[I]} \right\rangle = \frac{w_o^{[I]}}{L_o} \tag{14}$$

(5) *Interorganisational interdependencies* comprise formal and informal relations that emerge from the cross-organisation exchanges between individuals. Formal interorganisational degree ($w_{o_{int}}^{[F]}$) is the sum of weights of formal edges between individuals in organisation o and individuals in other organisations (L_r) in interorganisational network m, where O is the number of organisations in the interorganisational network m (15). Informal interorganisational degree ($w_{o_{int}}^{[I]}$) is the sum of weights of informal edges between individuals in organisation o and individuals in other organisations (L_r) in interorganisational network m (16). Interorganisational relations constitute an important part of interdependencies as they may facilitate interorganisational information exchange, knowledge sharing, innovation transfer, and support.

$$w_{o_{int}}^{[F]} = \sum_{i=1}^{L_o} \sum_{\substack{r=1 \\ r \neq o}}^{O} \sum_{j=1}^{L_r} w_{ij}^{[F]} \tag{15}$$

$$w_{o_{int}}^{[I]} = \sum_{i=1}^{L_o} \sum_{\substack{r=1 \\ r \neq o}}^{O} \sum_{j=1}^{L_r} w_{ij}^{[I]} \tag{16}$$

The average formal ($\langle w_{int}^{[F]} \rangle$) and informal ($\langle w_{int}^{[I]} \rangle$) interorganisational degrees ((17) and (18)) are calculated by normalization of the degrees by a total of individuals (N) in interorganisational network m.

$$\left\langle w_{int}^{[F]} \right\rangle = \frac{\sum_{o=1}^{O} w_{o_{int}}^{[F]}}{N} \tag{17}$$

$$\left\langle w_{int}^{[I]} \right\rangle = \frac{\sum_{o=1}^{O} w_{o_{int}}^{[I]}}{N} \tag{18}$$

4. Multilevel Interdependencies Model to Analyse Organisational Resilience

In this section, we bridge the multilevel conceptualisation of interdependencies with resilience analysis. More specifically, we present the application of conceptualisation to investigate the relation between interdependencies and organisational performance in normal and disrupted conditions. Based on the real world data, we construct an agent-based model of multilevel organisational interdependencies of two organisations. We demonstrate results of calculated interdependencies measures and performance simulations. The following research questions guided our analysis:

(1) What type of structure makes an organisational system more resilient?

(2) How does the degree of organisational interdependencies change at multiple levels?

(3) Does higher degree of organisational interdependencies contribute to better performance?

(4) Which of the interdependencies' levels contribute the most to the organisational performance?

4.1. Materials and Methods. We used the agent-based method and built a model of organisational interdependencies, which comprised formal and informal relations within and between two organisations. The model of interdependencies was constructed from the sociometric data gathered in August 2017. The data was collected at individual level (N=54) by roster questionnaires that were distributed in two collaborating organisations, which operate within the security services sector in Southeast Asia. Organisation A is a research, training, and operational support centre. Organisation B is an operational support centre to enhance well-being and operational effectiveness of another organisation's employee. Each of the organisations comprised 5 work units. The data described six multiplex social networks: reporting relationships, authority, acquaintance, friendship, familiarity, and problem-solving (Table 1).

Following the work of Luo [35], Haythornthwaite [36], Luo and Cheng [44], Krackhardt and Hanson [45], and Soda and Zaheer [92], the formal relations in the model consisted from reporting and authority networks and the informal interactions comprised acquaintance, friendship and familiar ties. For informal relations only mutual links (links that were confirmed by both connected agents) were considered as reliable and included in the model. The multiplex model was specified by the vector of the symmetric adjacency matrices (undirected graph) of formal and informal relations: A = $\{A^{[F]}, A^{[I]}\}$. Each of the matrices was constructed through the aggregation procedure that resulted in multiplex edge types, which allowed us to specify the simulation parameters (Appendix A). The pairs of agents are connected by either (1) both formal and informal edges, or (2) a formal edge, or (3) an informal edge (Figure 3).

The strength of social exchanges was reflected in the weight values given to each type of formal and informal edges. Qualitative and quantitative characteristics of the edges are described in Table 2. The formal and informal weight values range from 0 to 1. The weights of informal edges are based on the system of social relationships introduced by Luo [35] that governs complex social transactions dependently on proportions of instrumental and expressive exchange dimensions. According to Luo [35], Luo and Cheng [44] the strongest and most efficient resource exchanges are facilitated through familiarity ties ('rules of favour'), succeeded by friendship ('rules of need'), and acquaintance ('rules of equity'). The weights of formal edges are aligned with the hierarchical logic of formal employment relationships reflected in an organisational chart, which determine chains of command (authority) and formal information exchanges (reporting) [36]. In this light, the authority relations, representing formal vertical relations, are defined as the strongest, and are followed by reporting relationships that include both formal horizontal and vertical relations. The weights of aggregated formal and informal links have been sums of single network weight values (e.g. an edge representing single reporting and authority has a weight equal to 0.8 which is a sum of 0.2 and 0.6).

TABLE 1: Networks data used in the model.

	Social Network Questions	Relationships	Application in model
(1)	With whom do you like to discuss your daily work? [35]	Acquaintance (instrumental exchange)	Informal relations (trust-based)
(2)	With whom do you talk about your private affairs during your daily chats? [44]	Friendship (expressive exchange)	
(3)	Suppose that your colleague asks you to help his/her friend. Whose friends would you help? (adapted from Luo [35])	Familiar (instrumental and expressive exchange)	
(4)	To whom do you report about you work progress?	Reporting (instrumental exchange)	Formal relations
(5)	Who is your direct supervisor?	Authority (instrumental exchange)	
(6)	Whom do you ask for help when you encounter a work-related problem, for which you couldn't find a solution yourself?	Problem-solving (instrumental exchange)	Task demand
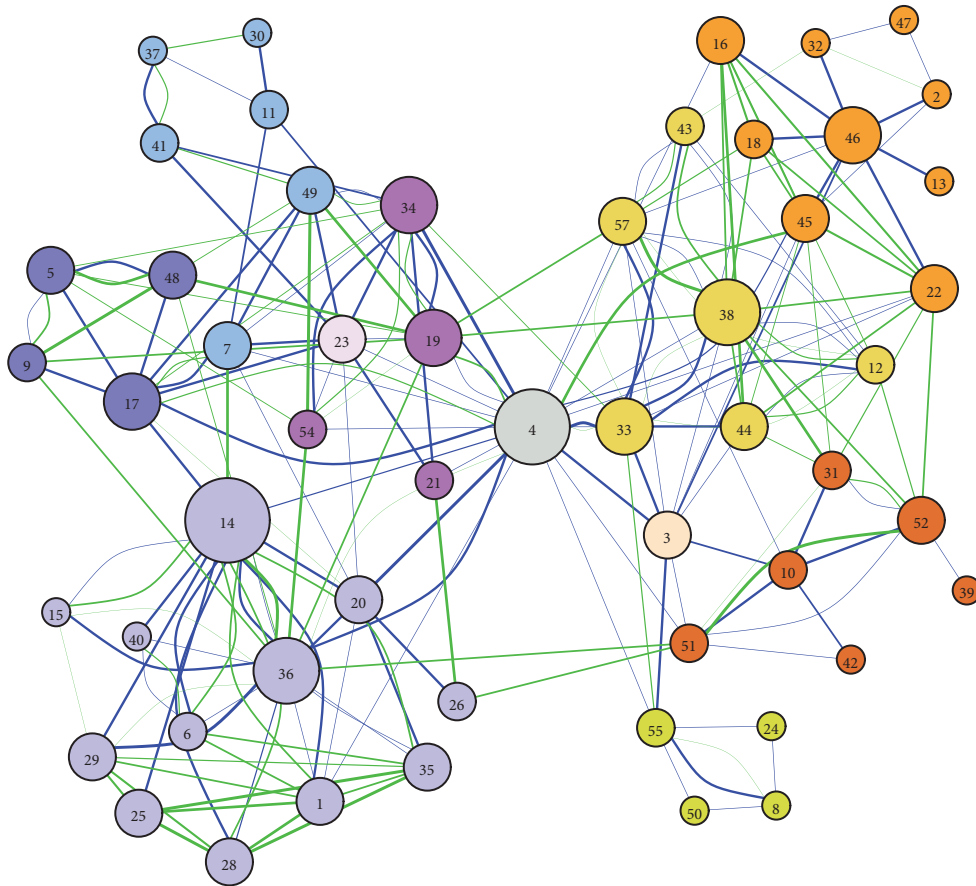
FIGURE 3: Multiplex model of organisational interdependencies. Nodes' colours represent work units. Node size depicts node strength. Blue and green edges are formal and informal relations. Individual no. 4 works in both organisations.

4.1.1. Simulations. To investigate how structure of organisational interdependencies affects organisational resilience, we used the model to conduct series of simulations both in normal and disrupted conditions. The simulations highlight two important aspects from the organisational systems functioning, i.e., task interdependence (agent must cooperate with other agents to complete a task) and adaptation through collective problem solving (the agent support each other directly and indirectly to solve a problem). In all simulations, more than 80% of all agents are given tasks to complete. All tasks are specified by *tasks demand* [T_D], which is a list of resources that agent needs to gather to complete the task.

TABLE 2: Quantitative and qualitative characteristics of existing relations.

Structure type	Edge types	Weight	Social exchange strength	Qualitative characteristics
Formal	One way reporting	0.2	Weak	(i) Weak instrumental exchange (ii) Limited reliability (iii) No expectation of reciprocation (iv) Occasional exchanges
	Mutual reporting	0.4	Moderate	(i) Moderate instrumental exchange (ii) Moderate reliability (iii) Expectation of reciprocation (iv) Occasional exchanges
	Authority	0.6	Strong	(i) Strong instrumental exchange (ii) High reliability (iii) No expectation of reciprocation (iv) Frequent exchanges
	Authority, One way reporting	0.8	Very strong	(i) Strong instrumental exchange (ii) Very high reliability (iii) No expectation of reciprocation (iv) Very frequent exchanges
	Authority, Mutual reporting	1.0	Extremely strong	(i) Strong instrumental exchange (ii) Extremely high reliability (iii) Expectation of reciprocation (iv) Very frequent exchanges
Informal	Acquaintance	0.1	Weak	(i) Weak instrumental and weak expressive exchange (ii) Moderate level of trust (iii) Rules of fair exchange (iv) Expectation of instant reciprocation (v) Limited reliability, often insufficient in obtaining valuable resources (vi) Occasional exchanges
	Friendship	0.3	Moderate	(i) Weak instrumental and strong expressive exchange (ii) High level of trust (iii) Rules of need (iv) Expectation for reciprocation, but not instant (v) High reliability (vi) Long-term, occasional exchanges (ad hoc when needed)
	Friendship, Acquaintance	0.4	Strong	(i) Moderate instrumental and strong expressive exchange (ii) High level of trust (iii) Expectation of reciprocation, but not instant (iv) High reliability (v) Long-term, occasional exchanges
	Familiarity	0.6	Very strong	(i) Strong instrumental and moderate expressive exchange (ii) High level of trust (iii) Rules of favour exchange (iv) Expectation of reciprocation, but not instant (v) Very high reliability (vi) Long-term and frequent exchanges (vii) Strong enough to be a bridge to connect to other agents
	Familiarity, Acquaintance, Friendship	>= 0.7	Extremely strong	(i) Strong instrumental and medium or strong expressive exchange (ii) High level of trust (iii) Expectation of reciprocation, but not instant (iv) Extremely high reliability (v) Long-term and very frequent exchanges (vi) Strong enough to be a bridge to connect to other agents

```
Input: tasks, other_agents, networks, disrupted_agents
Output: agent_task_completion_time
t = 0
(1) for task in tasks do
(2)    for other_agent in other_agents do
(3)       if task = other_agent[resource] then
(4)          if other_agent not in disrupted_list then
(5)             task = 100%
(6)             t+= transfer_time(other_agent, agent)
(7)          else
(8)             for subs_agent in other_agent[shared_resources] do
(9)                if subs_agent not in disrupted_list then
(10)                  if distance(subs_agent, agent)= 1 then    ▷1 step
(11)                     agent[task]+ = subs_agent[subs_resources]
(12)                     t+ = transfer_time(subs_agent, agent)
(13)                  else if distance(subs_agent, agent)= 2 then    ▷2 step
(14)                     if link(1) and link(2) fulfill the 2-step rule then
(15)                        agent[task]+ = subs_agent[subs_resources]
(16)                        t+ = transfer_time(subs_agent, agent)
(17) if all tasks ≥ 100 then
(18)    return t
(19) else
(20)    return nan
```

ALGORITHM 1: Task completion and resource sharing.

The task demand was generated from the directed problem-solving network (Table 1) through selecting resources that belonged to the agents connected by out-going edges (i.e., agents that an agent would contact if a work-related problem occurs). If the number of resources was more than five, only five resources were randomly selected, so the task demand lists ranged from one to maximum five resources. The task is completed when 100% of resources listed on the task demand are gathered (Algorithm 1). Each agent has a resource of his own (e.g., agent J has a resource J), and each network edge possesses a weight value that reflects strength of the relationship, as described in Table 2. The strength of relationship is reflected also in the parameter of speed and, thus, transfer time (Appendix A, Table 3). The transfer time needed to connect and pass the resource between the agents is depended on the edge type. For each agent's step, the transfer time is calculated by subtracting the sum of formal and informal edges' speed values from 100. In each simulation we calculate the total time based on all transfer times needed to accomplish tasks to measure performance of agents, units, organisations, and interorganisational network. In the resources search, the agent use both formal and informal structure in the maximum distance of two edges. Based on social exchange theory, the agent will choose always the fastest way. To ensure the realism of simulations, the speed priority is given to all formal types of links (the most time efficient), so the agent will choose and informal edge only if there is no formal connection. In line with work of Luo [35], the 2-step resource sharing is only possible when two conditions are fulfilled: (1) the first step edge is at least moderately strong informally (i.e., follows the rules of need or favour) or weak informally but the condition of

instant reciprocation is fulfilled (rules of equity); and (2) the second step edge contains the familiarity component (rules of favour). In this case, the total transfer time is sum of all steps needed to share the resource (see more details in Algorithm 1 and examples in Appendix B).

On the basis of the task demand lists, we create two types of *major work disruptions*:(1) targeted disruption (unavailability) of agents the most needed (i.e. agents with highest in-degree centrality in each unit; for bigger units at least 20% of agents were disrupted) to complete the task demands (N=9, 20% of agents with tasks) and (2) random disruption of agents, which had a resource needed to complete at least one task demand (N=9, the same number of agents as in targeted disruptions). While in normal conditions, an agent, to complete his task, looks for a resource directly from the original resource (an agent who owns it); in disrupted conditions, the agents are allowed to use the substituting resources if the original resource is not available anymore (owning agent disrupted) (see Algorithm 1). The substituting resource is conditioned by existence of the edge type that comprises exchange of work-related information, i.e., reporting (formal relations) and acquaintance (informal relations). To create substituting resources we considered directed networks of reporting and acquaintance (only mutual links), in which incoming edges granted an agent 40% and 20% of substituting resource respectively. Thus, in case of simultaneous reporting and acquaintance incoming links, the maximum value of substituting resource in question is 60% (Appendix A, Table 3). When we disrupt the organisational system, to complete his task, an agent first checks his available substituting resources and next looks for other agents who can substitute the needed resource to reach the needed amount (100%).

TABLE 3: Simulation parameters.

Relations	Aggregated edge type	Contained network edges	No. of existing multiplex links	Weight	Speed	Share of other agents resources in disruption condition (directional)	Contribution to 2-step resource sharing
Formal	Rep_one_way	One way reporting	55	0.2	40	0.4	.
	Rep_mutual	Mutual reporting	3	0.4	45	0.4	.
	Superv	Authority	6	0.6	50	.	.
	Superv_rep_one_way	Authority, One way reporting	52	0.8	55	0.4	.
	Superv_rep_mut	Authority, Mutual reporting	4	1.0	60	0.4	.
Informal	Acq	Acquaintance	15	0.1	5	0.2	Step 1 + T_D
	Friend	Friendship	12	0.3	10	.	Step 1
	Acq_friend	Friendship, Acquaintance	17	0.4	15	0.2	Step 1
	Famil	Familiarity	28	0.6	20	.	Steps 1 and 2
	Acq_famil	Familiarity, Acquaintance	6	0.7	25	0.2	Steps 1 and 2
	Friend_famil	Familiarity, Friendship	11	0.9	30	.	Steps 1 and 2
	Acq_friend_famil	Familiarity, Friendship, Acquaintance	8	1.0	35	0.2	Steps 1 and 2

4.2. Results. To answer what type of structure makes an organisational system more resilient, we examined the importance of formal and informal relations in determining the system behaviour under normal and disrupted conditions (Figure 4). We conduct simulations of systems including (1) only formal relations (blue line), (2) only informal relations (green line), and (3) both formal and informal relations, i.e., the overall network (red line). Figure 4 presents average values from simulations of normal conditions (n=1000), random disruptions (n=3000), and targeted disruptions (n=1000). We measure the system performance by number of completed tasks and time to complete the tasks. The simulations results showcase that organisational system with the overall structure (including formal and informal relations) performs better than the system, which has only formal structure or only informal structure. Importantly, the overall structure system performance is not a sum of formal and informal structures' performance. In this way, the overall system performance is not directly related to the number of existing links but rather their strength (relations quality) and other structural properties. Furthermore, it is crucial to acknowledge that in all conditions the formal structure plays a significant role in system performance and the organisational system that possesses only informal structure cannot perform well. In normal condition, all tasks are completed in both systems, but the system with overall structure completes the tasks faster. In case of both disruptions, the performance of both systems drops, as some tasks cannot be completed. However, it is very important to highlight that during the disturbed conditions, informal structure complements the formal structure (by 8%

and 15% in random and targeted disruptions respectively). More specifically, the informal structure supplements the unavailable formal connections, and this results in higher number of completed tasks as well as faster completion of the tasks. In sum, in general terms, the organisational system with both formal and informal relations performs better under normal and disrupted conditions, thus is more resilient.

In addition, to rule out the factor of different links number and ensure the correct interpretation of the results, we have sampled formal and informal structures to have the same number of edges (precisely 90). In this way, we constructed the new overall structure that comprised 90 formal and 90 informal edges. We revised the task demand lists in accordance with the agents existing in the new overall structure. The new structures were sampled five times and simulations for the three conditions were repeated (n=1000, n=3000, and n=1000). The average sampling results are presented in Figure 5. In general, the sampled results are consistent with the non-sampled (Figure 4). However, due to the executed changes that resulted in less efficient topology of new structures (smaller density, smaller transitivity, and longer average path; see Table 4), we reported performance drop of sampled formal and informal structures in normal condition and random disruption. The performance drop was bigger in case of formal than informal structure, especially in normal condition. That was due to higher deteriorative discrepancy between the sampled and the nonsampled formal structure in comparison to the sampled and non-sampled informal structure. In particular, the biggest disadvantages
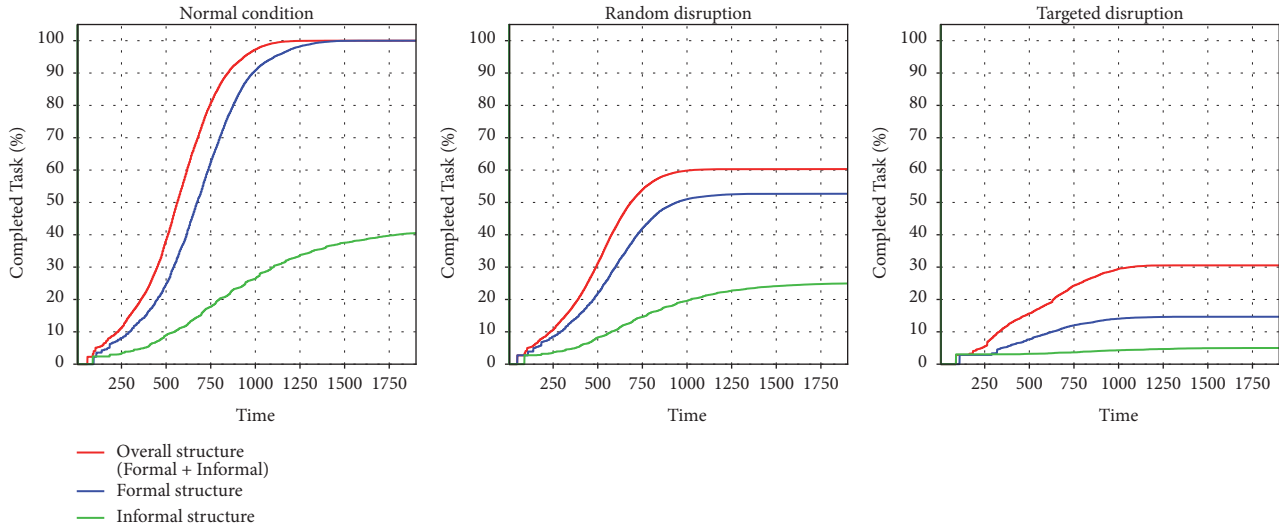
FIGURE 4: Performance of organisational structures in normal and disrupted conditions.
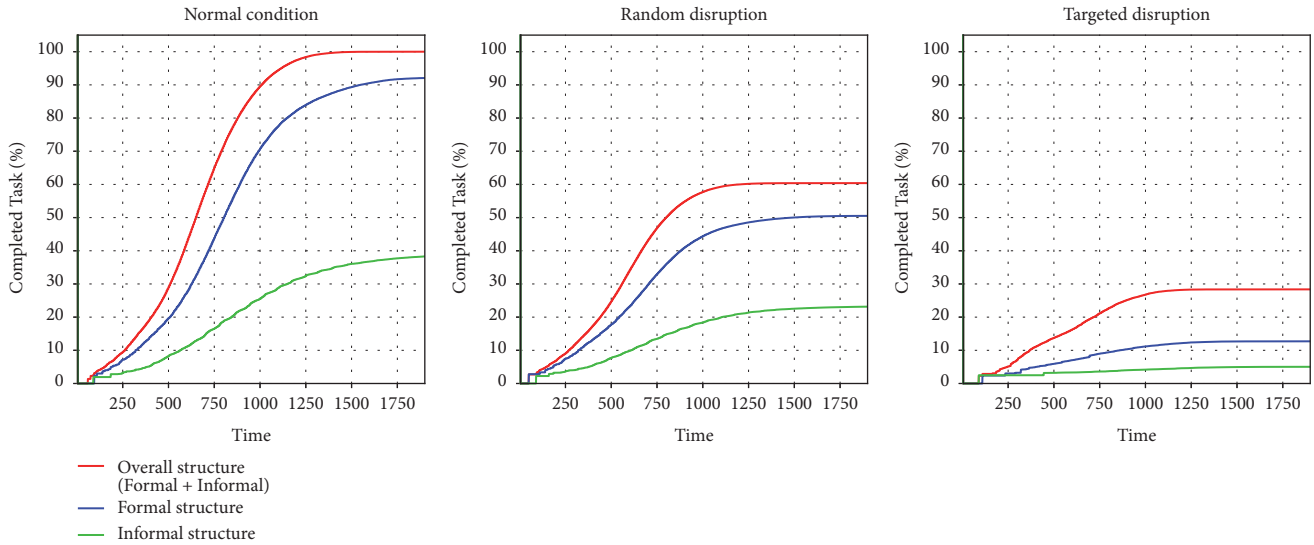
FIGURE 5: Average performance of sampled organisational structures in normal and disrupted conditions.

sampled formal structure concerned smaller transitivity, longer average path, and longer diameter, which highly influenced structure dynamics (see Table 4).

In relation to the formal and informal structure, we investigated which types of edges facilitated most of exchanges in normal and disrupted conditions structure contributes the most to the organisational performance (Figure 6). The usage of all types was relatively stable in all conditions. In total, most of exchanges were facilitated by both formal and informal edges (on average 46%) or only by formal edges (on average 43%). Approximately 11% of exchanges were conducted by informal edges. Similar pattern was observed over the simulation time. These results highlight with more details and reiterate the important contribution of the informal structure to the organisational system performance in normal and disrupted conditions.

Next, we examined the values of the formal and informal interdependencies degrees at multiple levels (Appendix C, Tables 5–9). Table 10 presents the summarised results of the degree analysis. The higher levels of the analysis, the higher were the values of non-averaged degrees due to the increasing number of considered network elements and their complex nature. The nonaveraged values of individual, intraunit, interunit, intraorganisational, and interorganisational degrees ranged from 0 to 72.8, and the averaged degrees ranged from 0 to 5.5. While both formal and informal individual, interunit, and intraorganisational degree values had moderate dispersion, the informal intraunit (SD = 7.8, mean = 5.3) and average formal interunit values (SD = 1.6, mean = 1.4) were characterised by high dispersion; that is, the degree values were very widely distributed. As data concern the relations within one interorganisational network, the
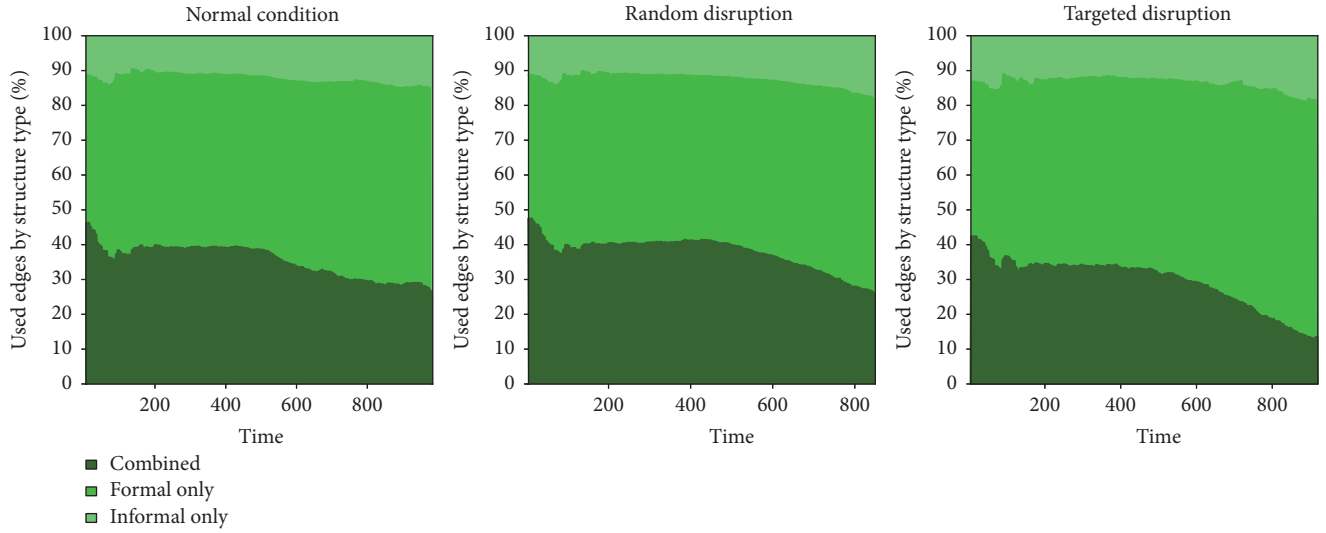
(a) Used edges by structure type over simulation time
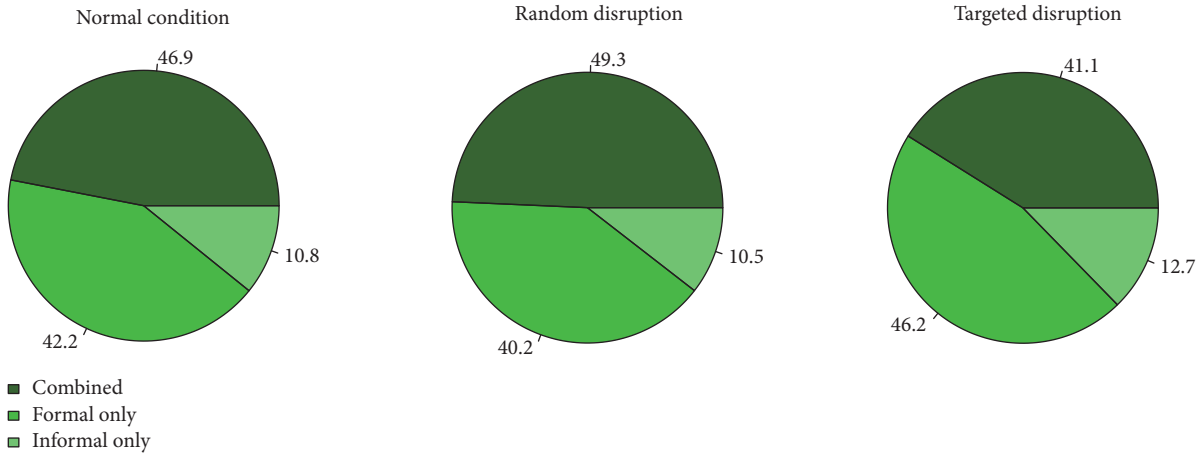
(b) Used edges by structure type in total

FIGURE 6: Types of edges used in normal and disrupted conditions.

TABLE 4: Properties of nonsampled and sampled structures.

Structure	No. nodes	No. edges	Density	Transitivity	Average path length	Diameter
Non-sampled overall	54	217	0.15	0.44	2.66	2.40
Non-sampled formal	54	120	0.08	0.33	2.94	3.00
Non-sampled informal	43	97	0.11	0.43	3.40	3.90
Average sampled overall	53	180	0.13	0.38	2.81	2.66
Average sampled formal	52	90	0.07	0.24	3.31	3.76
Average sampled informal	42	90	0.1	0.39	3.42	3.76

variation at the interorganisational-level interdependencies was not possible to assess.

The multilevel interdependence degrees are network centrality measures, i.e., rankings, which can be investigated by the Spearman's rank correlation coefficient ρ and the Kendall's rank correlation coefficient τ [93, 94]. Using the correlation coefficients we examined if higher degrees contribute to the better performance, that is, if there is a negative relation between formal and informal degrees and time to complete

a task at multiple levels (Table 11 and Figure 7). We adjust the significance values applying the sequential Bonferroni method [88] to avoid inflated risk of Type 1 error related to multiple comparisons. Subplots of Figure 7 are attached in the Supplementary Material for comprehensive image analysis. As each agent had a task that required different number of resources, each value of time to complete the task was normalized by number of demanded resources. In normal and random disruption conditions, there was a significant

Table 5: Individual degree.

Agent Id	Org.	Unit	Formal Individual Degree	Informal Individual Degree
1	A	A3	1.4	4.2
2	B	B1	1.2	0.1
3	B	B	4.6	0.0
4	A/B	A/B	9.4	2.1
5	A	A1	1.8	2.4
6	A	A3	1.2	2.2
7	A	A4	3.6	1.3
8	B	B4	1.2	0.1
9	A	A1	1.0	2.8
10	B	B2	3.8	0.0
11	A	A4	2.2	0.0
12	B	B3	1.6	1.7
13	B	B1	0.8	0.0
14	A	A3	8.0	4.9
15	A	A3	1.0	0.8
16	B	B1	1.0	3.9
17	A	A1	6.4	1.2
18	B	B1	0.8	2.9
19	A	A2	1.2	5.6
20	A	A3	4.8	0.1
21	A	A2	1.8	0.9
22	B	B1	1.2	4.2
23	A	A	5.6	0.0
24	B	B	0.4	0.0
25	A	A3	0.8	3.6
26	A	A3	0.8	1.5
28	A	A3	1.2	4.2
29	A	A3	1.8	2.6
30	A	A4	0.8	0.4
31	B	B2	1.0	2.4
32	B	B1	1.0	0.2
33	B	B3	6.0	1.1
34	A	A2	5.2	2.0
35	A	A3	1.2	4.2
36	A	A3	5.4	4.3
37	A	A4	1.0	0.8
38	B	B3	1.8	6.6
39	B	B2	0.2	0.0
40	A	A3	1.2	0.4
41	A	A4	2.2	0.8
42	B	B2	0.8	0.0
43	B	B3	1.4	1.1
44	B	B3	1.4	3.1
45	B	B1	1.6	3.5
46	B	B1	6.4	0.0
47	B	B1	0.4	0.0
48	A	A1	1.6	3.4
49	A	A4	2.6	2.9
50	B	B	0.4	0.0
51	B	B2	1.6	2.3

Table 5: Continued.

Agent Id	Org.	Unit	Formal Individual Degree	Informal Individual Degree
52	B	B2	1.4	3.0
54	A	A2	1.2	2.7
55	B	B4	2.2	0.5
57	B	B3	2.2	2.6

Table 6: Intraunit degree.

Unit	Formal Intraunit Degree	Informal Intraunit Degree	Number of Individuals	Average Formal Intraunit Degree	Average Informal Intraunit Degree
A	0.4	0.0	2	0.20	0.00
A1	6.8	5.0	4	1.70	1.25
A2	4.8	1.8	4	1.20	0.45
A3	25.2	27.4	12	2.10	2.28
A4	6.4	2.4	6	1.07	0.40
B	1.6	0.0	2	0.80	0.00
B1	12.4	8.2	9	1.38	0.91
B2	7.6	3.0	6	1.27	0.50
B3	10.8	7.6	6	1.80	1.27
B4	3.2	0.2	4	0.80	0.05

Table 7: Interunit degree.

Unit	Formal Interunit Degree	Informal Interunit Degree	Number of Individuals	Average Formal Interunit Degree	Average Informal Interunit Degree
A	11.0	1.1	2	5.50	0.55
A1	4.0	4.8	4	1.00	1.20
A2	4.6	7.9	4	1.15	1.97
A3	3.6	4.4	12	0.30	0.37
A4	6.0	3.8	6	1.00	0.63
B	6.6	1.0	2	3.30	0.50
B1	2.0	6.6	9	0.22	0.73
B2	1.2	3.5	6	0.20	0.58
B3	3.6	7.1	6	0.60	1.18
B4	1.0	0.4	4	0.25	0.10

Table 8: Intraorganisational degree.

Org.	Formal Organisational Degree	Informal Organisational Degree	Number of Individuals	Average Formal Organisational Degree	Average Informal Organisational Degree
A	72.8	58.6	28	2.60	2.09
B	50.0	37.6	27	1.85	1.39

Table 9: Interorganisational degree.

Interorg. network	Formal Interorg. Degree	Informal Interorg. Degree	Number of Individuals	Average Formal Interorg. Degree	Average Informal Interorg. Degree
m	0.0	2.7	54	0.00	0.05

(a) Individual level degrees

(b) Average intraunit level degrees

(c) Average interunit level degrees
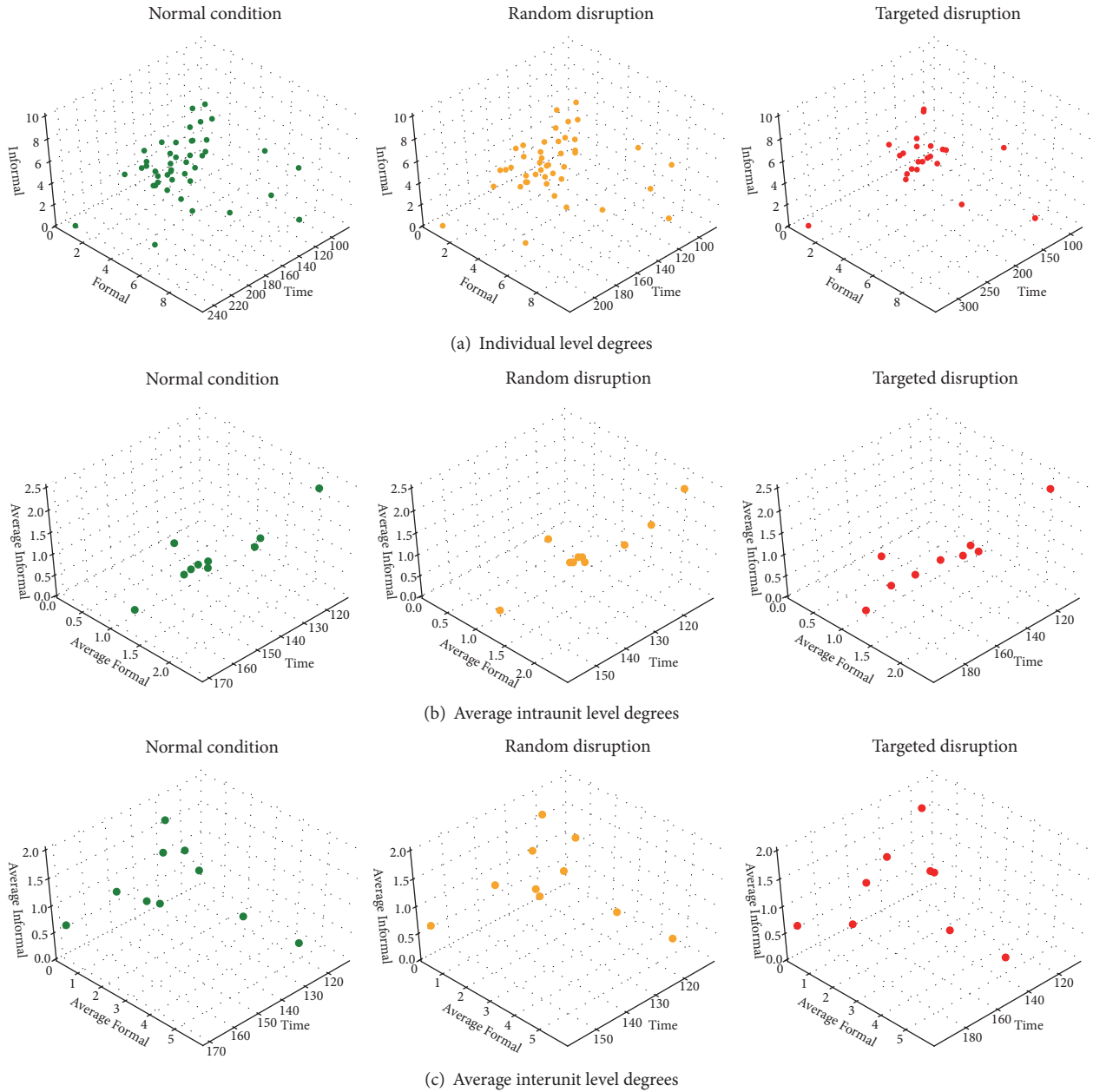
Figure 7: Relation between average interunit level degrees and task performance in normal and disrupted conditions.

negative correlation between informal individual degree and time to complete a task; i.e., the higher informal individual degrees (Figure 7(a)) the faster task completion ($\rho = -.35$, $\rho = -.33$). Also, agents with higher formal individual degree tended to complete tasks faster in targeted disruption ($\rho = -.26$).

At the intraunit level (Figure 7(b)), while in normal condition and random disruption there was no correlation between the formal and the informal intraunit degrees and time, in the targeted disruption both average formal and informal intraunit degrees were moderately correlated with better performance ($\rho = -.44$, $\rho = -.43$). At the interunit level (Figure 7(c)), in all conditions, units with higher average formal interunit degree performed better, i.e., needed shorter

time to complete tasks ($\rho = -.54$, $\rho = -.46$, and $\rho = -.43$) than units with lower degrees. At the intraorganisational level (Table 12), Organisation A with both higher average formal and informal intraorganisational degrees performed better than Organisation B, which had lower degrees. As the model concerns the only one interorganisational network, the relations between interorganisational degrees and performance were not possible to assess.

There are three edge levels, which contain unique information about the organisational interdependencies, i.e., intraunit, interunit, and interorganisational. We examined their usage to assess which of them contribute the most to the organisational performance (Figure 8). In total, most
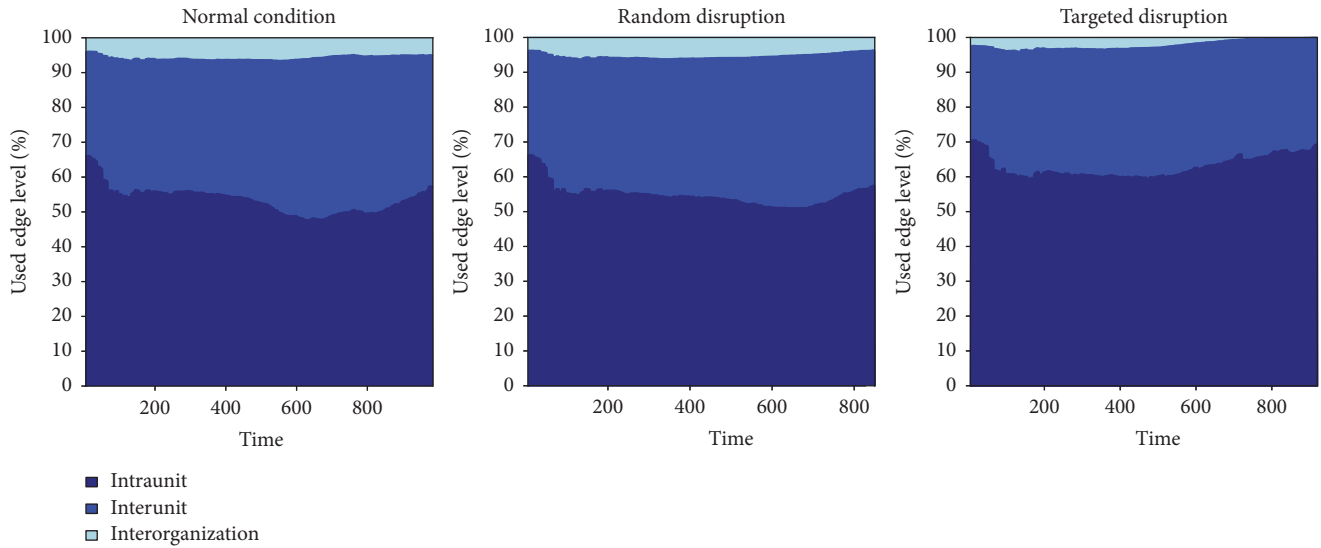
(a) Usage of edge levels over simulation time
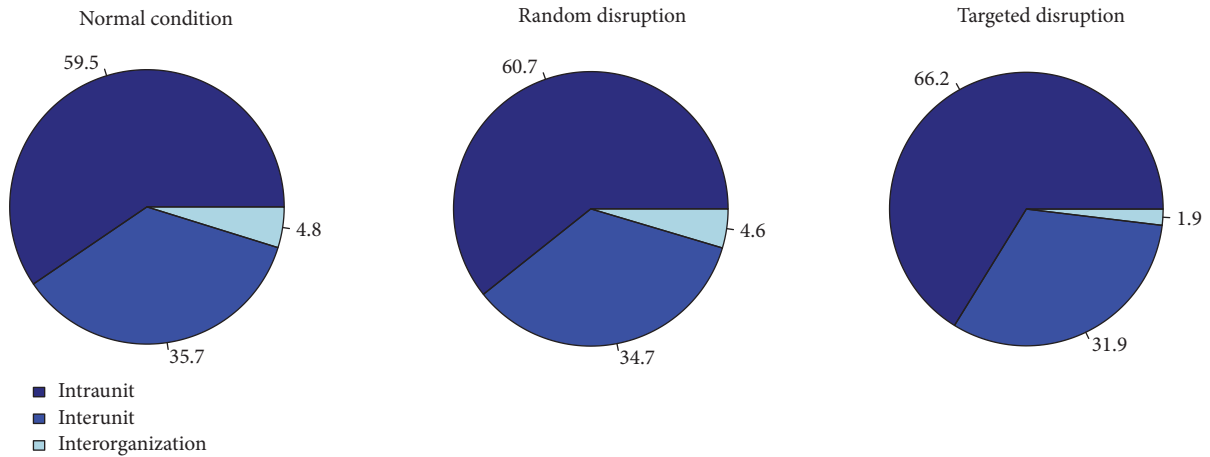
(b) Usage of edge levels in total

FIGURE 8: Levels of edges used in normal and disrupted conditions.

of the organisational exchanges were facilitated through the intraunit (on average 62%) and interunit (on average 34%) edges. The interorganisational edges constituted only 4% of the total number of used edges. While the usage of inter- and intraunit edges was stable in the normal and disrupted conditions, the usage of interorganisational edges decreased in the targeted disruption. This was due to the small number of interorganisational edges and low redundancy, which made this level prone to the disturbances caused in the targeted disruption.

5. Discussion and Implications

Organisational interdependence is a multidimensional construct that can be conceptualised at multiple levels. On one hand, the interdependencies make an organisational system more complex [30, 95]. On the other hand, the existence of multiplex relationships that comprise interdependencies is a natural feature of modern organisations [96, 97]. Because

system's safety is an emergent property [63] and partial analysis of the system can be misleading [84], it is crucial to have an in-depth understanding of the interdependencies' dynamics that happen at multiple levels. The analysis of structural properties of organisational interdependencies helps to identify patterns and assess needs and opportunities both in normal and disrupted conditions. By better understanding multilevel interdependencies we can reduce vulnerability and increase the ability to withstand dynamic changes and, thus, enhance organisational resilience.

Resilience is embedded in people's behaviour, and it is built by proactive approach to mobilizing resources, abilities to respond and perform under a variety of conditions. In this study, we view organisational resilience as a systemic property, which requires management of relational dynamics at multiple organisational levels. For that reason, the concept is highly relevant to resilience analysis in the organisational context. Our multilevel conceptualisation of interdependencies considered two dimensions of organisational

TABLE 10: Summary of interdependence degrees values at multiple levels.

Level	Interdependence Degree	N	SD	Mean	Max	Min
Individual	Formal individual	54	2.3	2.4	9.4	0.2
	Informal individual	54	1.7	1.9	6.6	0.0
Intraunit	Formal intraunit	10	6.8	7.9	25.2	0.4
	Informal intraunit	10	7.8	5.6	27.4	0.0
	Average formal intraunit	10	0.5	1.2	2.1	0.2
	Average informal intraunit	10	0.7	0.7	2.3	0.0
Interunit	Formal interunit	10	2.8	4.4	11.0	1.0
	Informal interunit	10	2.5	4.1	7.9	0.4
	Average formal interunit	10	1.6	1.4	5.5	0.2
	Average informal interunit	10	0.5	0.8	2.0	0.1
Intraorg.	Formal intraorganisational	2	11.4	61.4	72.8	50.0
	Informal intraorganisational	2	10.5	48.1	58.6	37.7
	Average formal intraorganisational	2	0.4	2.2	2.6	1.9
	Average informal intraorganisational	2	0.4	1.7	2.1	1.4
Interorg.	Formal interorganisational	1	0.0	0.0	0.0	0.0
	Informal interorganisational	1	0.0	2.7	2.7	2.7
	Average formal organisational degree	1	0.0	0.0	0.0	0.0
	Average informal organisational degree	1	0.0	0.05	0.05	0.05

TABLE 11: Correlation between interdependence degrees and time to complete task.

Degree	Time to complete task(s)		
	Normal condition	Random disruption	Targeted disruption
Formal individual	-.024 (.000)	-.098 (-.055)	-.262∗ (-.193∗)
Informal individual	**-.351∗∗(-.252∗∗)**	**-.325∗ (-.239∗)**	-.174 (-.134)
Average formal intraunit	.058 (.068)	-.095 (-.023)	-.442 (-.295)
Average informal intraunit	.073 (.091)	-.085 (.000)	-.427 (-.273)
Average formal interunit	-.537 (-.432∗)	-.463 (-.341)	-. 427 (-.250)
Average informal interunit	.450 (.270)	.316 (.180)	-. 377 (-.270)
			Spearman's ρ (Kendall's τ)

Bold correlation values are significant at α level corrected by sequential Bonferroni method [88].
∗ $p < 0.001$ level (1-tailed); ∗$p < 0.05$ level (1-tailed).

relations—formal and informal—and proposed measures to investigate their structure across levels. With well-mapped organisational interdependencies, we can examine how an organisational system behaves under normal and disturbed circumstances. The empirical study proved the structure of interdependencies influences the efficiency of organisational performance. In the model, in line with the state-of-art literature [7–9], we highlighted that task interdependence affects the dynamics and outcomes of organisational relationships. The results showed that an organisational system with a rich structure that combined both formal and informal relations performed better both in normal and disrupted conditions and thus was more resilient than the system based only on the formal relations. The formal structure appeared to meaningfully contribute to organisational performance. However, our results also underlined the importance of informal structure that substantially complements and substitutes the formal structure, especially in the disrupted conditions. We showcased that the trust-based relationships strongly affect agent's decision-making and flow of network resources. The

stronger informal relations, the faster transfer time and more resources are shared, as agents are more willing to provide assistance to each other. In addition, in most of simulated exchanges, the actors were connected by both formal and informal edge at the same time. These results confirm that well-established relationships (both formal and informal) within and between organisations condition organisational performance both in normal and disrupted conditions.

The analysis of introduced multilevel interdependence degrees can determine the organisation's ability to both work as expected, in normal conditions, as well as to create emergent response networks to confront the unexpected, mitigate consequences and adapt to the 'new normal'. The degree of organisational interdependence is not uniformed; rather it changes depending on the analysed relationship level. Facing the unexpected, task execution very often has to be changed and improvised due to unavailability of the needed resources. Individuals, units or organisations that normally are loosely coupled can be tightly coupled during a period of disruption. The interactions' changes happen at multiple

TABLE 12: Relation between average intraorganisational level degrees and task performance in normal and disrupted conditions.

Org.	Average formal intraorganisational degree	Average informal intraorganisational degree	Time to complete task(s)		
			Normal condition	Random disruption	Targeted disruption
A	2.60	2.09	126	118	122
B	1.85	1.36	145	134	157

levels; therefore, it is crucial to not only acknowledge the importance of individual interdependencies, but also higher interdependencies levels, such as intraunit, interunit, intraorganisational and interorganisational. Even though shaped by individuals' interactions, the higher-level degrees have unique features and provide new information that is crucial for resilience analysis of an organisational system. At multiple levels, the results showed there is a negative relation between the interdependence degrees, the time needed to complete a task. This was the most evident at the individual, interunit and organisational levels. Most of the simulated exchanges were facilitated through intra- or interunit edges, which at the same time contribute the most to the overall system performance. The analysis of the multilevel interdependence degrees indicated also that individuals, units, and organisations are more structurally embedded in the organisational system and, thus, potentially can benefit the most to organisational performance in normal and disrupted conditions and at the same time contribute to organisational resilience.

The results of this research imply that well-managed interdependencies are crucial to ensure resilient organisational performance. In the following lines, we present a few implications of this research. From our results, we can conclude that a resilient organisation should aim to have the dense and strong of organisational interdependencies, especially at intra- and interunit levels, which are constituted by the individuals' exchanges within the organisation. Some of the practices to reach this goal could include (1) decentralising (flattening) the organisational structure, (2) changing the organisational structure to the matrix model, which imposes higher number of formal interactions between units, and (3) nourishing organisational culture and providing environment that will help to activate formulation of new informal relations within and between units as well as strengthen the existing ones (organising social events, retreats, and common spaces). Furthermore, both formal and informal organisational structure contribute to organisational performance during normal and disrupted situations; thus both of them should be part of in-depth analysis while establishing crisis management plans, procedures and practices. A resilient organisation should have balanced amount of formal and informal interdependencies, which can facilitate complex relational exchanges when the performance conditions are both certain and uncertain. The formal structure is very efficient in enabling the collaborative work; however, it is prone to the disruptions, and often it is the informal structure that supplements it during the disruptions. It is important to stay aware of the differences and advantages of the two types of relations. The analysis of informal structure can help to unmask concealed organisational patterns, needs, and

opportunities. This information can be used to eliminate hidden organisational vulnerabilities and use overlooked potentials. In practice, the strengths of informal structure could be recognised by acknowledgment of the role of informal leaders in emergency preparedness activities (e.g., drills and exercises), emergency response plans, contingency plans, and the business continuity plan.

This article aimed to introduce a new way of conceptualising organisational interdependencies and present its usefulness for resilience analysis. Our analysis had limited focus on measuring the direct effects or causality between organisational interdependencies and organisational performance in expected and unexpected conditions. Future research should take up the challenge to broaden analysis of the impact of organisational interdependencies on system resilience. The described application of the multilevel perspective on organisational interdependencies gave new insights into the structure of organisational interdependencies embedded in the specific context, which shapes the perceptions and behaviours of the involved actors. Future studies should consider cross-disciplinary analyses of the organisational interdependencies in other organisational contexts and larger samples in order to build up the comprehensive of how the interdependencies are shaped in different organisational systems and how they affect system performance. These future analyses, along with the preliminary results described in this paper, should be the insightful base of next practices for resilient design and management of organisational systems.

6. Conclusion

Organisations are complex interdependent systems, which require management efforts to remain resilient. As noted above, we argue for correlation between resilience and organisational interdependencies. Interdependencies in organisations have been studied from various perspectives; however, most of the analyses were conducted only at the single level and overlooked the multilevel character of interdependencies. In response to this gap, we proposed a multilevel approach to better comprehend the complexity of interdependencies in organisational systems. With this paper we contributed to the study of interdependencies and resilience by introducing a multilevel conceptualisation of interdependencies in organisational systems and presenting its application for resilience analysis. Adapting the most plausible definition of interdependencies as exchange relationships, our paper sheds light on how those relationships are shaped across multiple organisational levels and suggests how they could be decoupled in order to handle their complexity. We used the system and multilevel theories

T_D: [F, L]
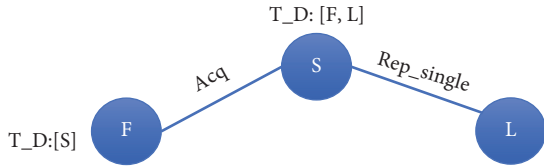
FIGURE 9: One-step resource sharing in normal conditions.
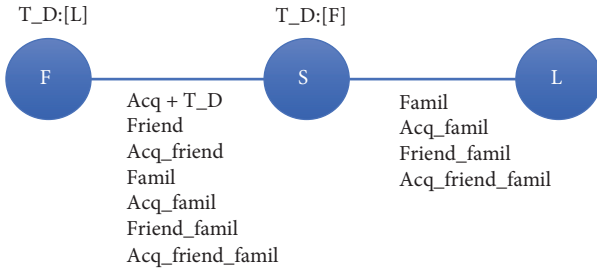
FIGURE 10: Two-step resource sharing in normal conditions.

to explore five organisational levels, including individual, intraunit, interunit, intraorganisational, and interorganisational. We argued that interdependencies are a discontinuous phenomenon across levels that does not express uniform pattern. Accordingly, we employed the configural compilation approach to describe interdependencies' features and proposed their measures at multiple levels. In addition to formal relationships, our conceptualisation underlines the significance of informal trust-based relationships, a notion that provides a new insight on the origins of interdependence. Furthermore, we applied the multilevel interdependencies conceptualisation into the analysis of organisational resilience and presented the relations between interdependencies and organisational performance at multiple levels in normal and disrupted conditions. Finally, we discussed how managing multilevel interdependencies is crucial to reduce vulnerability and to build up, maintain, and enhance resilience of organisational systems. By introducing the multilevel conceptualisation, we hope to pave a preliminary path to managing the complexity of the interdependencies in organisational systems. At the same time, we advance the analysis of the multilevel relationships between interdependencies and resilience as a promising step towards improving organisational design and resilience management.

Appendix

A.

See Table 3.

B.

B.1. 1-Step Resource Sharing in Normal Condition (See Figure 9)

(i) Agent F needs a resource of agent S (T_D:[S]): agent S gives resource to agent F because they are connected

by an acquaintance edge (acq) and agent S has agent F on his task demand list (T_D:[F]); Transfer time = 100 − 5 = 95.

(ii) Agent S needs a resource of agent L (T_D:[L]): agent L gives resource to agent S because they are connected by a single-way reporting edge; Transfer time: 100 − 40= 60.

B.2. 2-Step Resource Sharing in Normal Condition (See Figure 10)

Scenario. Agent F needs a resource of agent L (T_D : [L]); as they are not directly connected, he asks agent S for help.

Step 1. If agent S and agent F are connected by an instrumental tie, i.e., acquaintance (acq), agent S will facilitate resource sharing only if agent's F resource is on his task demand list (T_D), that is, when the fair exchange can take place. In case of other informal links the resource connection will be directly facilitated.

Step 2. Agent S can ask agent L for a favour to help agent F only if they are connected by strong instrumental and expressive ties, i.e., containing familiarity component (famil, acq_famil, friend_famil, and acq_friend_famil)

B.3. Disruption (See Figure 11)

Scenario. Agent F's task demand is resource J (agent J). Agent J is disrupted (not available).

(1) Agent F has initial 20% of needed resource of agent J due to the informal connection (acq_friend_famil) comprising exchange of work-related information (component of 'acq').

(2) As agent J reports to agent S, agent F asks agent S for additional 40% of agent J (superv_rep_one_way). There is also a parallel informal link (friend) which facilitates faster transfer; Transfer time = 100 − (55+10) = 35.

(3) Because agent F and agent S share strong expressive ties (friends) and agent S and H share strong expressive and instrumental tie (famil), agent S asks agent H for a favour to help with missing resource part (40%); Transfer time = (100 − (55+10)) + (100-20) = 35 + 80= 115.

(4) Total transfer time: 150.

C.

See Tables 1–A.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.
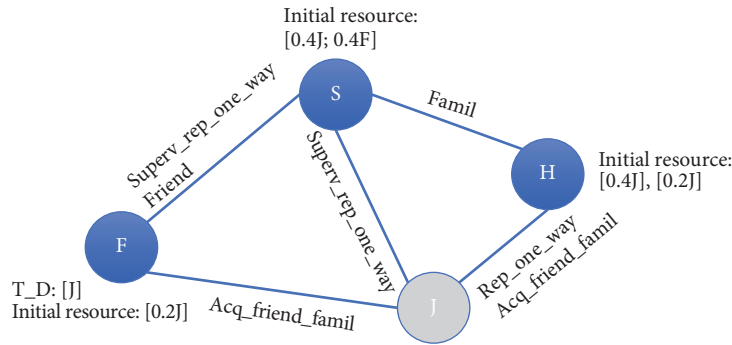
Figure 11: Resource sharing in disruption conditions.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. M. Rousseau, "Issues of level in organizational research: Multi-level and cross-level perspectives," *Research in Organizational Behavior*, vol. 7, no. 1, pp. 1–37, 1985.

[2] R. House, D. M. Rousseau, and M. Thomas-Hunt, "The Meso paradigm: a framework for the integration of micro and macro organizational behavior," in *Research in Organizational Behavior*, L. L. Cummings and B. M. Staw, Eds., Jai Press, Greenwich, Conn, USA, 1995.

[3] M. A. Hitt, P. W. Beamish, S. E. Jackson, and J. E. Mathieu, "Building theoretical and empirical bridges across levels: Multilevel research in management," *Academy of Management Journal (AMJ)*, vol. 50, no. 6, pp. 1385–1399, 2007.

[4] T. R. La Porte, *Organized Social Complexity: Challenge to Politics and Policy*, vol. 71, Princeton University Press, Princeton, NJ, USA, 1975.

[5] B. P. Cohen and R. Arechavala-Vargas, *Interdependence, Interaction, and Productivity*, 1987.

[6] H. C. Dekker, "Control of inter-organizational relationships: Evidence on appropriation concerns and coordination requirements," *Accounting, Organizations and Society*, vol. 29, no. 1, pp. 27–49, 2004.

[7] R. C. Liden, S. J. Wayne, and L. K. Bradway, "Task interdependence as a moderator of the relation between group control and performance," *Human Relations*, vol. 50, no. 2, pp. 169–181, 1997.

[8] D. Tjosvold, S. Sasaki, and J. W. Moy, "Developing commitment in Japanese organizations in Hong Kong: Interdependence, interaction, relationship, and productivity," *Small Group Research*, vol. 29, no. 5, pp. 560–581, 1998.

[9] G. Van der Vegt, B. Emans, and E. Van de Vliert, "Motivating effects of task and outcome interdependence in work teams," *Group & Organization Management*, vol. 23, no. 2, pp. 124–143, 1998.

[10] M. Kilduff and D. Krackhardt, *Interpersonal Networks in Organizations: Cognition, Personality, Dynamics, and Culture*, Cambridge University Press, New York, NY, USA, 2008.

[11] H. Oh, M.-H. O. Chung, and G. Labianca, "Group social capital and group effectiveness: The role of informal socializing ties," *Academy of Management Journal (AMJ)*, vol. 47, no. 6, pp. 860–875, 2004.

[12] R. Gulati, "Alliances and networks," *Strategic Management Journal*, vol. 19, no. 4, pp. 293–317, 1998.

[13] K. G. Provan, A. Fish, and J. Sydow, "Interorganizational networks at the network nevel: A review of the empirical literature on whole networks," *Journal of Management*, vol. 33, no. 3, pp. 479–516, 2007.

[14] C. Tomkins, "Interdependencies, trust and information in relationships, alliances and networks," *Accounting, Organizations and Society*, vol. 26, no. 2, pp. 161–191, 2001.

[15] A. V. Shipilov, "Firm scope experience, historic multimarket contact with partners, centrality, and the relationship between structural holes and performance," *Organization Science*, vol. 20, no. 1, pp. 85–106, 2009.

[16] B. R. Barringer and J. S. Harrison, "Walking a tightrope: creating value through interorganizational relationships," *Journal of Management*, vol. 26, no. 3, pp. 367–403, 2000.

[17] C. Alter and J. Hage, *Organizations Working Together*, vol. 15, Sage Publications, Newbury Park, Calif, USA, 1993.

[18] M. Aiken and J. Hage, "Organizational interdependence and intra-organizational structure," *American Sociological Review*, pp. 912–930, 1968.

[19] K. H. Roberts, "New challenges in organizational research: High reliability organizations," *Organization & Environment*, vol. 3, no. 2, pp. 111–125, 1989.

[20] J. D. Thompson, *Organizations in Action: Social Science Bases of Administrative Theory*, Transaction Publishers, New Brunswick, Canada, London, 1967.

[21] D. J. Brass, "Being in the right place: a structural analysis of individual influence in an organization," *Administrative Science Quarterly*, pp. 518–539, 1984.

[22] R. S. Burt, "A note on cooptation and definitions of constraint," *Social Structure and Network Analysis*, pp. 219–233, 1982.

[23] M. Granovetter, "Economic action and social structure: the problem of embeddedness," *American Journal of Sociology*, vol. 91, no. 3, pp. 481–510, 1985.

[24] M. Granovetter, "Problems of explanation in economic sociology," *Networks and Organizations: Structure, Form, and Action*, vol. 25, p. 56, 1992.

[25] H. A. Simon, "The organization of complex systems," in *Hierarchy Theory*, H. H. Pattee, Ed., Braziller, New York, NY, USA, 1973.

[26] L. Bertalanffy, *General System Theory : Foundations, Development, Applications*, Braziller, New York, NY, USA, 1969.

[27] D. Katz and R. L. Kahn, "Organizations and the system concept," *Classics of Organization Theory*, pp. 161–172, 1978.

[28] J. G. Miller, *Living Systems*, McGraw-Hill, New York, NY, USA, 1978.

[29] T. R. La Porte, "High reliability organizations: Unlikely, demanding and at risk," *Journal of Contingencies and Crisis Management*, vol. 4, no. 2, pp. 60–71, 1996.

[30] V. Milch and K. Laumann, "Interorganizational complexity and organizational accident risk: A literature review," *Safety Science*, vol. 82, pp. 9–17, 2016.

[31] S. W. J. Kozlowski and K. J. Klein, "A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes," in *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*, K. J. Klein and S. W. J. Kozlowski, Eds., pp. 3–90, Jossey-Bass, San Francisco, Calif, USA, 2000.

[32] S. P. Borgatti and P. C. Foster, "The network paradigm in organizational research: A review and typology," *Journal of Management*, vol. 29, no. 6, pp. 991–1013, 2003.

[33] M. Kilduff and W. Tsai, *Social Networks and Organizations*, Sage, Thousand Oaks, Calif, USA, 2003.

[34] E. Lazega and T. A. Snijders, *Multilevel Network Analysis for the Social Sciences: Theory, Methods and Applications*, vol. 12, Springer, 2016.

[35] J.-D. Luo, "Guanxi revisited: an exploratory study of familiar ties in a chinese workplace," *Management and Organization Review*, vol. 7, no. 2, pp. 329–351, 2011.

[36] C. Haythornthwaite, "Social network analysis: An approach and technique for the study of information exchange," *Library & information science research*, vol. 18, no. 4, pp. 323–342, 1996.

[37] D. J. Brass, J. Galaskiewicz, H. R. Greve, and W. Tsai, "Taking stock of networks and organizations: a multilevel perspective," *Academy of Management Journal (AMJ)*, vol. 47, no. 6, pp. 795–817, 2004.

[38] T. P. Moliterno and D. M. Mahony, "Network theory of organization: A multilevel approach," *Journal of Management*, vol. 37, no. 2, pp. 443–467, 2011.

[39] R. Gulati, "Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances," *Academy of Management Journal*, vol. 38, no. 1, pp. 85–112, 1995.

[40] J. Coleman, *Foundations of Social Theory*, Belknap Press of Harvard University Press, Mass, USA, 1990.

[41] B. Kogut, "The network as knowledge: Generative rules and the emergence of structure," *Strategic Management Journal*, vol. 21, no. 3, pp. 405–425, 2000.

[42] E. Ostrom and J. Walker, *Trust and reciprocity: Interdisciplinary lessons for experimental research*, Russell Sage Foundation, New York, NY, USA, 2003.

[43] A. Mehra, M. Kilduff, and D. J. Brass, "The social networks of high and low self-monitors: Implications for workplace performance," *Administrative Science Quarterly*, vol. 46, no. 1, pp. 121–146, 2001.

[44] J.-D. Luo and M.-Y. Cheng, "Guanxi circles' effect on organizational trust: bringing power and vertical social exchanges into intraorganizational network analysis," *American Behavioral Scientist*, vol. 59, no. 8, pp. 1024–1037, 2015.

[45] D. Krackhardt and J. R. Hanson, "Informal networks: the company behind the chart," *Harvard Business Review*, vol. 71, no. 4, pp. 104–111, 1993.

[46] J. M. Flach, J. S. Carroll, M. J. Dainoff, and W. I. Hamilton, "Striving for safety: communicating and deciding in sociotechnical systems," *Ergonomics*, vol. 58, no. 4, pp. 615–634, 2015.

[47] D. Krackhardt and R. N. Stern, "Informal networks and organizational crises: an experimental simulation," *Social Psychology Quarterly*, pp. 123–140, 1988.

[48] G. I. Rochlin, "Informal organizational networking as a crisis-avoidance strategy: US naval flight operations as a case study," *Organization & Environment*, vol. 3, no. 2, pp. 159–176, 1989.

[49] E. E. Umphress, G. Labianca, D. J. Brass, E. Kass, and L. Scholten, "The role of instrumental and expressive social ties in employees' perceptions of organizational justice," *Organization Science*, vol. 14, no. 6, pp. 738–756, 2003.

[50] Y.-A. De Montjoye, A. Stopczynski, E. Shmueli, A. Pentland, and S. Lehmann, "The strength of the strongest ties in collaborative problem solving," *Scientific Reports*, vol. 4, p. 5277, 2014.

[51] C. Folke, "Resilience: The emergence of a perspective for social-ecological systems analyses," *Global Environmental Change*, vol. 16, no. 3, pp. 253–267, 2006.

[52] C. Folke, S. R. Carpenter, B. Walker, M. Scheffer, T. Chapin, and J. Rockström, "Resilience thinking: Integrating resilience, adaptability and transformability," *Ecology and Society*, vol. 15, no. 4, 2010.

[53] G. C. Gallopín, "Linkages between vulnerability, resilience, and adaptive capacity," *Global Environmental Change*, vol. 16, no. 3, pp. 293–303, 2006.

[54] W. N. Adger, "Social and ecological resilience: are they related?" *Progress in Human Geography*, vol. 24, no. 3, pp. 347–364, 2000.

[55] A. D. Ong, C. S. Bergeman, T. L. Bisconti, and K. A. Wallace, "Psychological resilience, positive emotions, and successful adaptation to stress in later life," *Journal of Personality and Social Psychology*, vol. 91, no. 4, p. 730, 2006.

[56] Y. Sheffi and J. B. Rice Jr., "A supply chain view of the resilient enterprise," *MIT Sloan Management Review*, vol. 47, no. 1, p. 41, 2005.

[57] G. Hamel and L. Välikangas, "The quest for resilience," *Harvard Business Review*, vol. 81, no. 9, pp. 52–63, 2003.

[58] D. Paton and D. Johnston, *Disaster Resilience: An Integrated Approach*, Charles C Thomas Publisher, Springfield, Ill, USA, 2017.

[59] J. Tasic and S. Amir, "Informational capital and disaster resilience: the case of Jalin Merapi," *Disaster Prevention and Management*, vol. 25, no. 3, pp. 395–411, 2016.

[60] S. Amir and V. Kant, "Sociotechnical resilience: a preliminary concept," *Risk Analysis*, vol. 38, no. 1, pp. 8–16, 2018.

[61] V. Kant and J. Tasic, "Mapping sociotechnical resilience," in *The Sociotechnical Constitution of Resilience*, S. Amir, Ed., pp. 67–90, Springer, 2018.

[62] E. Hollnagel, D. D. Woods, and N. Leveson, *Resilience Engineering: Concepts and Precepts*, Ashgate Publishing, Aldershot, UK, 2006.

[63] N. Leveson, "A new accident model for engineering safer systems," *Safety Science*, vol. 42, no. 4, pp. 237–270, 2004.

[64] K. E. Weick and K. M. Sutcliffe, *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*, John Wiley & Sons, NJ, USA, 2011.

[65] K. E. Weick, K. M. Sutcliffe, and D. Obstfeld, "Organizing for high reliability: Processes of collective mindfulness," *Research in Organisational Behavior*, vol. 1, pp. 81–123, 1999.

[66] R. Bhamra, S. Dani, and K. Burnard, "Resilience: The concept, a literature review and future directions," *International Journal of Production Research*, vol. 49, no. 18, pp. 5375–5393, 2011.

[67] A. Boin and M. J. G. van Eeten, "The resilient organization," *Public Management Review*, vol. 15, no. 3, pp. 429–445, 2013.

[68] H. R. Heinimann and K. Hatfield, "Infrastructure resilience assessment, management and governance–state and perspectives," in *Resilience and Risk*, pp. 147–187, Springer, 2017.

[69] E. Hollnagel, *Safety-II in Practice: Developing the Resilience Potentials*, Taylor & Francis, 2017.

[70] K. Burnard and R. Bhamra, "Organisational resilience: Development of a conceptual framework for organisational responses," *International Journal of Production Research*, vol. 49, no. 18, pp. 5581–5599, 2011.

[71] G. S. Van Der Vegt, P. Essens, M. Wahlström, and G. George, "Managing risk and resilience," *Academy of Management*, vol. 58, no. 4, 2015.

[72] L. K. Comfort, Y. Sungu, D. Johnson, and M. Dunn, "Complex systems in crisis: anticipation and resilience in dynamic environments," *Journal of Contingencies and Crisis Management*, vol. 9, no. 3, pp. 144–158, 2001.

[73] L. Riolli and V. Savicki, "Information system organizational resilience," *Omega*, vol. 31, no. 3, pp. 227–233, 2003.

[74] K. M. Sutcliffe and T. J. Vogus, "Organizing for resilience," *Positive Organizational Scholarship*, pp. 94–110, 2003.

[75] W. Klibi, A. Martel, and A. Guitouni, "The design of robust value-creating supply chain networks: a critical review," *European Journal of Operational Research*, vol. 203, no. 2, pp. 283–293, 2010.

[76] M. K. Linnenluecke, "Resilience in business and management research: a review of influential publications and a research agenda," *International Journal of Management Reviews*, vol. 19, no. 1, pp. 4–30, 2017.

[77] N. Kapucu, "Interagency communication networks during emergencies: Boundary spanners in multiagency coordination," *The American Review of Public Administration*, vol. 36, no. 2, pp. 207–225, 2006.

[78] C. Jones, W. S. Hesterly, and S. P. Borgatti, "A general theory of network governance: Exchange conditions and social mechanisms," *Academy of Management Review (AMR)*, vol. 22, no. 4, pp. 911–945, 1997.

[79] L. Poppo and T. Zenger, "Do formal contracts and relational governance function as substitutes or complements?" *Strategic Management Journal*, vol. 23, no. 8, pp. 707–725, 2002.

[80] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, "Not so different after all: A cross-discipline view of trust," *Academy of Management Review (AMR)*, vol. 23, no. 3, pp. 393–404, 1998.

[81] K. J. Dooley, "A complex adaptive systems model of organization change," *Nonlinear Dynamics, Psychology, and Life Sciences*, vol. 1, no. 1, pp. 69–97, 1997.

[82] J. Sutherland and W.-J. Van Den Heuvel, "Enterprise application integration and complex adaptive systems," *Communications of the ACM*, vol. 45, no. 10, pp. 59–64, 2002.

[83] A. D. Meyer, A. S. Tsui, and C. R. Hinings, "Configurational approaches to organizational analysis," *Academy of Management Journal (AMJ)*, vol. 36, no. 6, pp. 1175–1195, 1993.

[84] F. Harary and M. F. Batell, "What is a system?" *Social Networks*, vol. 3, no. 1, pp. 29–40, 1981.

[85] G. Cowan, D. Pines, and D. Meltzer, *Complexity: Metaphors, Models, and Reality*, Addison-Wesley, Reading , Mass, USA, 1994.

[86] M. A. Griffin, "Interaction between individuals and situations: Using HLM procedures to estimate reciprocal relationships," *Journal of Management*, vol. 23, no. 6, pp. 759–773, 1997.

[87] F. P. Morgeson and D. A. Hofmann, "The structure and function of collective constructs: implications for multilevel research and theory development," *Academy of Management Review (AMR)*, vol. 24, no. 2, pp. 249–265, 1999.

[88] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, pp. 65–70, 1979.

[89] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.

[90] R. L. Cross and A. Parker, *The Hidden Power of Social Networks : Understanding How Work Really Gets Done in Organizations*, Harvard Business School Press, Boston, Mass, USA, 2004.

[91] R. Gulati, "Social structure and alliance formation patterns: A longitudinal analysis," *Administrative Science Quarterly*, pp. 619–652, 1995.

[92] G. Soda and A. Zaheer, "A network perspective on organizational architecture: Performance effects of the interplay of formal and informal organization," *Strategic Management Journal*, vol. 33, no. 6, pp. 751–771, 2012.

[93] L. Solá, M. Romance, R. Criado, J. Flores, A. García del Amo, and S. Boccaletti, "Eigenvector centrality of nodes in multiplex networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 23, no. 3, article no. 033131, 2013.

[94] V. Nicosia and V. Latora, "Measuring and modeling correlations in multiplex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 92, no. 3, article no. 032805, 2015.

[95] J. A. Rijpma, "Tight–coupling and reliability: connecting normal accidents theory and high reliability theory," *Journal of Contingencies and Crisis Management*, vol. 5, no. 1, p. 15, 1997.

[96] S. Ferriani, F. Fonti, and R. Corrado, "The social and economic bases of network multiplexity: Exploring the emergence of multiplex ties," *Strategic Organization*, vol. 11, no. 1, pp. 7–34, 2013.

[97] A. Shipilov, "Strategic multiplexity," *Strategic Organization*, vol. 10, no. 3, pp. 215–222, 2012.

Research Article

A Semantic Community Detection Algorithm Based on Quantizing Progress

Xu Han [iD],[1] Deyun Chen [iD],[1,2] and Hailu Yang [iD] [1,2]

[1]*School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, Heilongjiang 150080, China*
[2]*Postdoctoral Research Station of Computer Science and Technology, Harbin University of Science and Technology, Harbin, Heilongjiang 150080, China*

Correspondence should be addressed to Deyun Chen; chendeyun@hrbust.edu.cn and Hailu Yang; yanghailu@hrbust.edu.cn

The semantic social network is a kind of network that contains enormous nodes and complex semantic information, and the traditional community detection algorithms could not give the ideal cogent communities instead. To solve the issue of detecting semantic social network, we present a clustering community detection algorithm based on the PSO-LDA model. As the semantic model is LDA model, we use the Gibbs sampling method that can make quantitative parameters map from semantic information to semantic space. Then, we present a PSO strategy with the semantic relation to solve the overlapping community detection. Finally, we establish semantic modularity (SimQ) for evaluating the detected semantic communities. The validity and feasibility of the PSO-LDA model and the semantic modularity are verified by experimental analysis.

1. Introduction

With the development of society and the improvement of science and technology, semantic social networks are rapidly developed and many semantic networks, like Twitter and Weibo, have made an insignificant impact in our life so far. In these networks, different individuals have different small social "worlds" which are called communities [1]. Thus, researchers focus attention on community detection not only to divide networks into modules but also to make a deep insight into understanding interesting properties within the semantic social network. In practical application, semantic communities have a great promotion on intelligent information retrieval, marketing management, individual service, and other information management domains [2]. Heretofore, the research on community detection mainly reflects on the following three categories: topological community detection [3], community detection on overlapping construction [4], and semantic community detection.

The topological community detection represents the pioneer work, the goal of which is studying the topological constructions and dividing the social networks into several separate networks. The representative algorithms contain Modular Optimization [5], GN [6], and FN [7]. Then, researchers gradually focus on overlapping communities which can be more real than previous research networks. Therefore, CPM [8] was proposed to detect the overlapping communities. Soon afterwards, community detection on overlapping construction received more attention in social networks and many representative algorithms were proposed, including LFM [9], EAGLE [10], COPRA [11], DEMON [12], and so forth. Neuman and Yair [13] proposed an agglomerative spectral clustering method with conductance and edge weights. In their method, the most similar nodes are agglomerated based on eigenvector space and edge weights. But this method only is suitable for the nonsemantic social networks. Then, with the big interest in semantic network, semantic community detection came into researchers' eyes. Yang and McAuley [14] proposed the CESNA model to develop communities by using edge structure and node attributes. This method leads to more accurate community detection as well as improved robustness in the presence of noise in the network structure. But when this method applies into semantic network, it performs instable. Reihanian and

Ali [15] proposed a generic framework for overlapping community detection in social networks with special focus on rating-based social networks. This framework considers the information shared by the users in order to find meaningful communities. The most important feature of semantic communities is that the nodes in these communities not only have topological relationships, but also own semantic context. For the semantic data mining must be considered on the text analysis, and many semantic community detection algorithms applied the Latent Dirichlet Allocation (LDA) [16] model as the core model.

In the last few years, the analysis in semantic social network has become popular. Most of these algorithms utilize LDA model as the basic model. The SVM-DTW method proposed by Solera, Calderara, and Cucchiara [17] can work on the hierachical networks. This method makes simple structure and needs less input parameters, but the semantic context is not considered and the detected community has less connection with the real semantic network. Li and Ming and She [18] proposed the GRTM model which not only simulates users' interests as latent variables through their information, but also considers the connections between users as a result of their information. This method combines the context analysis with topological analysis and the similarity of the detected community is nearly close to the real semantic social network, but it is lack in the feature of sampling that would make some fuzzy irrelevant community. Xiao and Liu [19] proposed the GLDA-FP model which can be extended using the prediscretizing method which can help LDA model detect the topic evolution automatically, but the calculation required is large. As for the LCTA model proposed by Yin, Cao, and Gu [20] which makes the different topic distributions in different communities to make the model reasonable, this method has high accuracy in the result, but the number of communities needs to be preset and some hidden parameters need to be set up with experience.

In this paper, we propose a novel community detection algorithm for the objective of dividing nodes into clusters. The main characteristic of communities detected by this algorithm is that members of the same community have common or similar interests. We take into account the topic and keywords information in text from individuals' words through LDA model, then quantize semantic nodes, and map them into semantic space. Then, we get ideal virtual social communities after using Particle Swarm Optimization algorithm. Last but not least, we build a novel modular model and use the new function *SimQ* to evaluate the virtual social communities we make.

Compared with other models in semantic social network, such as lovain method model [21] and stochastic block model [22], the LDA model provides the probabilistic method so as to promote the foundation of mathematics. Then considering the following sampling, the Gibbs sampling can give an accurate and powerful mathematical proof for the convergence and solution of the LDA model, which is impossible to happen in the other semantic models. Combined with the PSO algorithm, the probability function compiled by LDA model can be closely integrated with the inertia weight and the constriction factor of the particles [23]. In performance measure, we propose a new module detecting evaluation model based on semantic information using the cosine function, which enriches the classic semantic detecting evaluation model.

The rest of the paper is organized as follows: Section 2 introduces LDA model in semantic network. Section 3 shows gibbs sampling and the proposed algorithm. In order to verify our approach, we conducted extensive experiments on a real data set. Performance evaluation and experimental results are shown and discussed in Sections 4 and 5. Finally, in Section 6 we make conclusions and envision further work.

2. Preliminaries

2.1. Community Detection Process. The problem of community detection belongs to NP-hard areas [24] which need initialize solutions at the beginning and optimize solutions constantly in the way of getting the best satisfying solution. The main goal of detecting semantic community is to form communities that individuals share common interests and probably they have similar characteristic [25]. So we show a novel idea that we focus on textual data of individuals' words. According to the complexity of community detection, we utilize the probabilistic graphical model–LDA to design network. This model has a most clearly hierarchical structure [26], and the scale of parameter spatial has no connection with the number of training documents.

First, we select topics and words from individuals' semantic information through LDA model. Then, we map semantic nodes into semantic space via Gibbs sampling method [27]. Last, in order to get more accurate communities, we use Particle Swarm Optimization (PSO) algorithm to form semantic communities. The proposed community detection algorithm is clearly explained in the following steps.

2.1.1. Similar Semantic Information Discovery. Every individual says different words as each node has its own information contents in semantic social network [28]. So we abstract semantic context into topic, and then we extract keywords from topic. Through semantic information, we convey some distributions to constrain our mess context [29]. In this way, dividing communities in semantic social network based on similar documents, topics, and keywords from social semantic contents make communities real [30]. The LDA probability model is shown in Figure 1.

In this section, we research LDA model on information contents. The relevant mathematical symbols for illustrating the LDA model are given in Table 1. LDA model assumes the following generative process for each node:

(1) $\theta \sim Dirichlet(\alpha)$. The parameter θ, which pertains to topic distribution, is subject to the Dirichlet distribution over a priori parameter α.

(2) $\varphi \sim Dirichlet(\beta)$. The parameter φ, which pertains to keyword distribution, is subject to the Dirichlet distribution over a priori parameter β.

(3) $z_i \mid \theta^{(d_i)} \sim Multinomial(\theta^{(d_i)})$. The topic z_i is subject to the multinomial distribution in case of topic distribution probability $\theta^{(d_i)}$.

TABLE 1: The symbol description.

SYMBOL	DESCRIPTION
N	Number of keywords in semantic social network
ω	Set of keywords in semantic social network, ω_i is the $i-th$ keyword in ω
d	Node set corresponding to keywords set ω, d_i is the $i-th$ node in the semantic social network
z	Topic set corresponding to keywords set ω, z_i is the $i-th$ topic in semantic social network
$\theta^{(d_i)}$	Topic distribution probability vector θ over node d_i
$\varphi^{(y)}$	Keyword distribution probability vector of topic y, $\varphi_{\omega_i}^{(y)}$ meaning the probability of keyword ω_i specific to topic y, $\varphi_{\omega_i}^{(y)} = P(\omega_i \mid z_i = y)$
α	A priori parameter over topic distribution probability specific to each node
β	A priori parameter over keyword distribution probability specific to a special topic
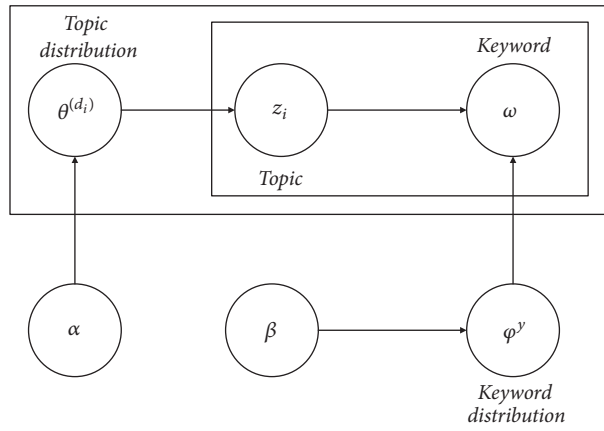
FIGURE 1: LDA probability model.

(4) $\omega_i \mid z_i, \varphi^{(z_i)} \sim Multinominal(\varphi^{(z_i)})$. The keyword ω_i is subject to the multinomial distribution in case of keyword distribution probability $\varphi^{(z_i)}$ over topic z_i.

The process of forming LDA model is shown in Algorithm 1. And M means the number of documents in the process.

3. Gibbs Sampling and PSO Strategy

3.1. Gibbs Sampling. Gibbs sampling [31] is a simple case of Markov-chain Monte Carlo (MCMC) [32] and aims at extracting a set of approximate samples from Markov-chain that is targeted to make a suitable probability distribution for converging to optimal solutions in high-dimensional models [33] such as LDA. According to the feature of Markov-chain, the probability-distribution function becomes the key to Gibbs sampling [34]. As for LDA in this text, we only sample topics in semantic social network; that is, we only need to consider hidden variety z_i. We denote $z_{\neg i}$ (topic set besides z_i) and $\omega_{\neg i}$ (set of keywords besides ω_i) to draw a posterior probability $P(z_i = y \mid z_{\neg i}, \omega_i)$. As for i, we can find the corresponding keyword ω_i. So the probability can be described as in the following equation.

$$P(z_i = y \mid z_{\neg i}, \omega_i) \propto P(z_i = y, \omega_i = t \mid \omega_{\neg i}, z_{\neg i}) \quad (1)$$

When $z_i = y$ and $\omega_i = t$ (t is one of the keywords in ω; y, which corresponds to t, is one of the topics in z), the probability $P(z_i = y, \omega_i = t \mid \omega_{\neg i}, z_{\neg i})$ only involves conjugate distribution of $d-th$ the document and $k-th$ topic under the Dirichlet-multinomial model.

We make $n_p^{[k]}$ as the number of $k-th$ topics in $d-th$ document, and the multinomial distribution can be described as

$$n_p = \left(n_p^{[1]}, n_p^{[2]}, \ldots n_p^{[K]} \right) \quad (2)$$

The number of $m-th$ keywords in $k-th$ topic, named $n_q^{[m]}$, can be shown as follows under multinomial distribution.

$$n_q = \left(n_q^{[1]}, n_q^{[2]}, \ldots n_q^{[M]} \right) \quad (3)$$

The posterior distribution of $\theta^{(d_i)}$ and $\varphi^{(z_i)}$ can be obtained in the following equations.

$$P\left(\theta^{(d_i)} \mid \omega_{\neg i}, z_{\neg i} \right) = Dirchlet\left(\theta^{(d_i)} \mid n_{p,\neg i} + \alpha \right) \quad (4)$$

$$P\left(\varphi^{(z_i)} \mid \omega_{\neg i}, z_{\neg i} \right) = Dirchlet\left(\varphi^{(z_i)} \mid n_{q,\neg i} + \beta \right) \quad (5)$$

$n_{p,\neg i}$ is the number of topics and $n_{q,\neg i}$ is the number of keywords.

```
(1)  Extract the keyword distribution, and φ ~ Dirichlet(β);
(2)  for each m ∈ [1, M] do
(3)      extract N keywords, and N ~ Poisson(φ);
(4)      Extract topic distribution, and θ ~ Dirichlet(α);
(5)      for each n ∈ [1, N] do
(6)          Extract a topic, and this topic obeys zᵢ ~ Multinomial(θ^(dᵢ));
(7)          Extract a keyword, and this keyword obeys ωᵢ ~ Multino min al (φ^(zᵢ));
(8)      end for
(9)  end for
```

ALGORITHM 1: The generative process of LDA.

The distribution probability $P(z_i = y, \omega_i = t \mid \omega_{\neg i}, z_{\neg i})$ can be calculated by (6)~(11).

$$P(z_i = y, \omega_i = t \mid \omega_{\neg i}, z_{\neg i}) = \int P(z_i = y, \omega_i = t, \varphi^{(z_i)},$$
$$\theta^{(d_i)} \mid \omega_{\neg i}, z_{\neg i}) d\theta^{(d_i)} d\varphi^{(z_i)} \tag{6}$$

$$= \int P(z_i = y, \theta^{(d_i)} \mid \omega_{\neg i}, z_{\neg i}) P(\omega_i = t, \varphi^{(z_i)} \mid \omega_{\neg i},$$
$$z_{\neg i}) d\theta^{(d_i)} d\varphi^{(z_i)} \tag{7}$$

$$= \int P(z_i = y \mid \theta^{(d_i)}) Dirichlet(\theta^{(d_i)} \mid n_{p,\neg i}$$
$$+ \alpha) d\theta^{(d_i)} \tag{8}$$

$$\cdot \int P(\omega_i = t \mid \varphi^{(z_i)}) Dirichlet(\varphi^{(z_i)} \mid n_{q,\neg i} + \beta) d\varphi^{(z_i)} \tag{9}$$

$$= \frac{n_{p,\neg i}^y + \alpha}{\sum_{f=1}^K n_{p,\neg i}^f + \alpha} \frac{n_{q,\neg i}^t + \beta}{\sum_{g=1}^V n_{q,\neg i}^g + \beta} \tag{10}$$

$$\Longrightarrow P(z_i = y \mid z_{\neg i}, \omega_i) \propto \frac{n_{p,\neg i}^y + \alpha}{\sum_{f=1}^K n_{p,\neg i}^f + \alpha}$$
$$\cdot \frac{n_{q,\neg i}^t + \beta}{\sum_{g=1}^V n_{q,\neg i}^g + \beta} \tag{11}$$

$n_{p,\neg i}^y$ is the amount of topics while $z_i = y$, $\sum_{f=1}^K n_{p,\neg i}^f$ is the amount of topics, $n_{q,\neg i}^t$ is the amount of keywords while $\omega_i = t$, and $\sum_{g=1}^V n_{q,\neg i}^g$ is the amount of keywords.

3.2. PSO Class Dependent LDA (PSO-LDA).

Particle Swarm Optimization (PSO) is an intelligent optimization algorithm. It was first proposed by J.Kennedy and R.C.Eberhart [35]. PSO algorithm has the advantages of simplified, rather quick convergence [36] speed and less controlling parameter, and so forth.

Compared with other optimization algorithms, such as Genetic Algorithm (GA), Ant Colony Optimization (ACO), and Simulate Anneal (SA), PSO algorithm has two attractive features: firstly, PSO optimizes the solution from the local optimum first and runs fast, which makes the algorithm more adaptable to the evolution of networks; secondly, particles in PSO can be mapped to nodes in semantic network; the process of finding the optimal solution in PSO is consistent with the birth process of the semantic community.

PSO puts a set of random solutions at system startup time and uses iterative search to find out optimal solutions [37]. In PSO, a solution of each optimization problem is called "particle". Each particle owns fitness value of itself. So we design a heuristic method to detect communities based on PSO. Each particle searches for the optimal solution by sharing social information between individuals.

In PSO-LDA, some LDA semantic feature is put into PSO. We use nodes in semantic social network mapping to "particle" in PSO and utilize semantic information vector of each node mapping to velocity of each particle in PSO. As for fitness value, we use information similar function instead. In PSO, we normalize that the nodes in semantic social network simulate the behavior of a "bird flock", where social sharing of information takes place, individuals' gains from the discoveries and previous experience of all other nodes during the search for food [38]. Thus, each node, called particle, in semantic social network which is called swarm, is assumed to "fly" over the search place looking for promising regions on the landscape.

First, we assume the search place is D−dimension space; and the $i − th$ particle position of the swarm is denoted as D−dimension, the vector $W_i = (w_{i1}, w_{i2}, \cdots, w_{id}, \cdots, w_{iD})$. Each particle has two pieces of message in the process: its "best" position with the smallest value (i.e., its personal best position) $P_i = (p_{i1}, p_{i2}, \cdots, p_{id}, \cdots, p_{iD})$ and the best function value of global particles in swarm (i.e., the global best position of all particles) $P_g = (p_{g1}, p_{g2}, \cdots, p_{gd}, \cdots, p_{gD})$. At each iteration, $i−th$ particle of the swarm updates its position and the velocity $V_i = (v_{i1}, v_{i2}, \cdots, v_{id}, \cdots, v_{iD})$ according to the following equation:

$$v_{id}^{s+1} = \eta v_{id}^s + \lambda_1 r_1^s (p_{id}^s - w_{id}^s) + \lambda_2 r_2^s (p_{gd}^s - w_{id}^s) \tag{12}$$

s is the current iteration, $j \in [1, 2, \cdots, D]$, $i \in [1, 2, \cdots, N]$, N represents the size of population, D is the dimension of the search place, η is the inertia weight, and λ_1 and λ_2 are two positive constants. r_1 and r_2 are study factors, that is, two random numbers extracted from the range [0, 1] for each dimension.

Input:
  The semantic social network gragh disposed by LDA;
Output:
  Useful transformable probability matrix;
  Step 0. Initialize proper parameters, inertia weight $\eta = 0.632$, constriction factor $\xi = 0.729$, study
  factors $r_1 = 2.8, r_2 = 1.3$, population size(the size of network) $M = 200$, particle size (the number of
  nodes in semantic social network) $N = 1000$ and maximum iteration $MI = 200$.
  Step 1. Initialize all particles and let $s = 0$;
  Step 2. Evaluate fitness of each particle;
  Step 3. Judge whether the ultimate criteria is satisfied. If $s > MI$, stop and jump to **Final**.; otherwise
  refresh variables according to the following steps;
  Step 4. Refresh p_{id} by comparing the current fitness of each particle with its own historical best position
   p_{id}, if p_{id} gets smaller, then change it with the current position;
  Step 5. Refresh p_{gd} by comparing the current best fitness of all particles with the historical best
  position p_{gd} of the whole swarm, if p_{gd} gets smaller, then change it with the current best position;
  Step 6. Refresh v_{id}^{s+1} and w_{id}^{s+1} using Eq (12) and Eq (13);
  Step 7. $s = s + 1$, return **Step 2**;
  Final.

ALGORITHM 2: Optimization algorithm by PSO.

In the search place, once velocity v_{id}^{s+1} updated, the $i - th$ particle position w_{id} is changed as in the following equation.

$$w_{id}^{s+1} = w_{id}^s + \xi v_{id}^{s+1} \tag{13}$$

ξ is a constriction factor which manages and regulates the velocity's magnitude to maintain a balance between exploration and exploitation and it can be calculated as follows:

$$\xi = \frac{2}{\left|2 - \lambda - \sqrt{\lambda^2 - 4\lambda}\right|} \tag{14}$$

$\lambda = \lambda_1 + \lambda_2, \lambda > 4$. The constriction factor has influence on the proposed algorithm; we discuss the issue in part 4. The pseudocode for PSO is described in Algorithm 2 [39].

4. Performance Measure

Generally speaking, the performance measure of semantic social network is mostly based on the topological construction. And the *EQ* model proposed by Shen et al. [40] is widely used in evaluating overlapping communities, which is described in the following equation:

$$EQ = \frac{1}{R} \sum_i \sum_{v \in C_i, w \in C_i} \frac{1}{O_v O_w} \left[A_{vw} - \frac{k_v k_w}{R} \right] \tag{15}$$

k_v is the degree of node v and k_w is the degree of node w, $R = \sum_{vw} A_{vw}$ is the total degree of the network, A_{vw} is the element of adjacency matrix of the network, O_v is the number of communities which the node v belongs to and O_w is the number of communities which the node w belongs to, and C_i is the $i - th$ community in the network. For we use both topological construction and semantic context to detect communities, a novel evaluation model named *SimQ*, which

we add information similarity into topological evaluation index, is given by the following equation.

$$SimQ$$
$$= \frac{1}{R_1} \sum_{i,j} \sum_{d_i \in C_i, d_j \in C_j} \frac{Sim\left(d_i, d_j\right)}{O_{d_i} O_{d_j}} \left[A_{d_i d_j} - \frac{k_{d_i} k_{d_j}}{R_1} \right] \tag{16}$$

d_i is the $i - th$ node and d_j is the $j - th$ node, O_{d_i} is the number of communities that the node d_i pertains and O_{d_j} is the number of communities that the node d_j pertains, $R_1 = \sum_{d_i d_j} A_{d_i d_j}$ is the total degree of the network, $A_{d_i d_j}$ is the element of adjacency matrix of the network, and the range of value for $SimQ$ is $(0, 1)$. As for the information similarity $Sim(d_i, d_j)$, we give a normal social graph $G = (D, E, X_{d_i/d_j}^K, Sim(d_i, d_j))$, where D is a set of nodes in the network and d_i/d_j is the $i/j - th$ node; E is the set of edges linking to graph nodes. The actual point of $Sim(d_i, d_j)$ is to measure the structural correlation of nodes and add semantic correlation components at the same time. This is more suitable for the basic characteristics of the semantic communities. Each node d_i has connection with an information vector $X_{d_i}^K = (x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)})$; $Sim(d_i, d_j)$ is the information similarity of two neighbor nodes i and j which is calculated as

$$Sim\left(d_i, d_j\right) = \frac{\sum_{i,j=1}^n \left(X_{d_i}^K X_{d_j}^K\right)}{\sqrt{\left(\sum_{i=1}^n \left(X_{d_i}^K\right)^2\right)\left(\sum_{j=1}^n \left(X_{d_j}^K\right)^2\right)}} \tag{17}$$

K is the dimension of the social network. In our method, if the semantic components of two nodes are close, the projection angles of these two nodes in two-dimensional space will be relatively small. On the contrary, the projection vectors are in contradictory situation.
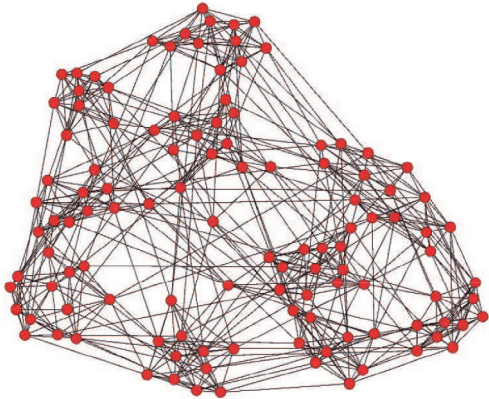
FIGURE 2: The graph of football network.
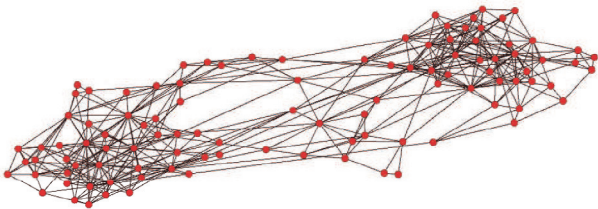
FIGURE 3: The graph of polbooks network.

5. Experimental Results

In this part, we would present and discuss the experiments with topics number analysis, evaluation criterion, real datasets, and different community detection algorithms, based on three datasets (the American College Football network dataset, the Krebs polbooks network dataset, and the dolphins network dataset).

5.1. The Analysis on Topics Number. The number of topics T, which is one of the input parameters in PSO-LDA model, can influence the compactedness of communities. So we choose the following three datasets to verify the effect of topics T over the result: (1) The American College Football network is shown in Figure 2. This network, created by Newman, is a complex social network about American College Football league. Nodes are regarded as football teams and one edge, between two neighbor nodes, represents that two football teams have played a match. It contains 115 nodes and 616 edges. (2) The Krebs polbooks network established by V.Kreb is shown in Figure 3. The nodes represent the politics books sold on Amazon. Generally, the books on political tendency are approximately divided into three classes. So in order to get topic distribution, Newman collected the political tendency in 3 steps away around each node. (3) The dolphins network collected by Newman is shown in Figure 4. The dolphins network is made up of two families, including 62 nodes and 159 edges. We simulate each node with the semantic information to fit on Dirichlet distribution.
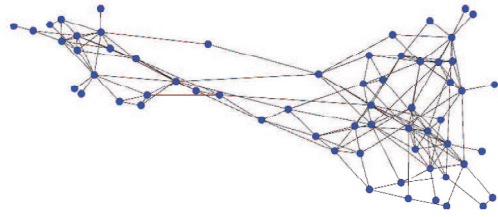
FIGURE 4: The dolphins network.

In this section, we use the topic number to experimentalize on three datasets (football, polbooks, and dolphins). Figure 5 shows the comparison of *EQ* and *SimQ* on the three datasets with $T = (1, 2, \cdots, 20)$. While the topic number T grows bigger and the topic distribution rises higher, the number of detected communities gets bigger as T rises. In Figure 5, when the topic number gets larger to a certain degree, the topic distribution tends to be stable, resulting in the increment of communities. From the comparison of *EQ* and *SimQ*, these two performance measure models tend to decrease as T increases, since the topic number T arrives at an optimal point. The optimal value of T is 6 in Figure 5.

For the sake of getting communities more intuitive, Figure 6 shows the detected communities of three datasets when T is 6, 12, and 18.

5.2. The Comparison on Different Optimization Algorithms. In this section, we do the comparison on different optimization algorithms with three network datasets above (dolphins, polbooks, and football). We compare the number of communities, the size of communities, runtime, and semantic concentration with PSO algorithm, Genetic Algorithm (GA), Ant Colony Optimization (ACO), and Simulate Anneal (SA). The result is shown in Figure 7. From Figure 7, we can see PSO algorithm makes more numbers of communities and smaller size of communities than others. As for runtime in PSO algorithm, it runs a little better than ACO and SA. The semantic concentration (*SC*) [41] is a function for measuring and testing degree of coagulation on specific topic and *SC* is shown in the following equation:

$$SC = \frac{\sum_{ij} Sim\left(d_i, d_j\right) \cdot \delta_{ij}}{\sum_{ij} Sim\left(d_i, d_j\right)} \tag{18}$$

δ_{ij} is the performance measure of communities links, while $\delta_{ij} = 1$ and only if i and j belong to the same community, there is a link between i and j. Compared with similarity function *SimQ*, *SC* makes focus on the stability of social groups in local environment. But what needs to be noted is that higher *SimQ* does not mean higher *SC* in communities and higher *SC* does not mean we can get the best divisions; this is because the overlapping part of communities can effect the semantic cohesion. So the ideal community construction should be suitable with *SimQ* and *SC*, and this also fits the performance measure of overall optimization and local optimization. Compared with GA, ACO, and SA
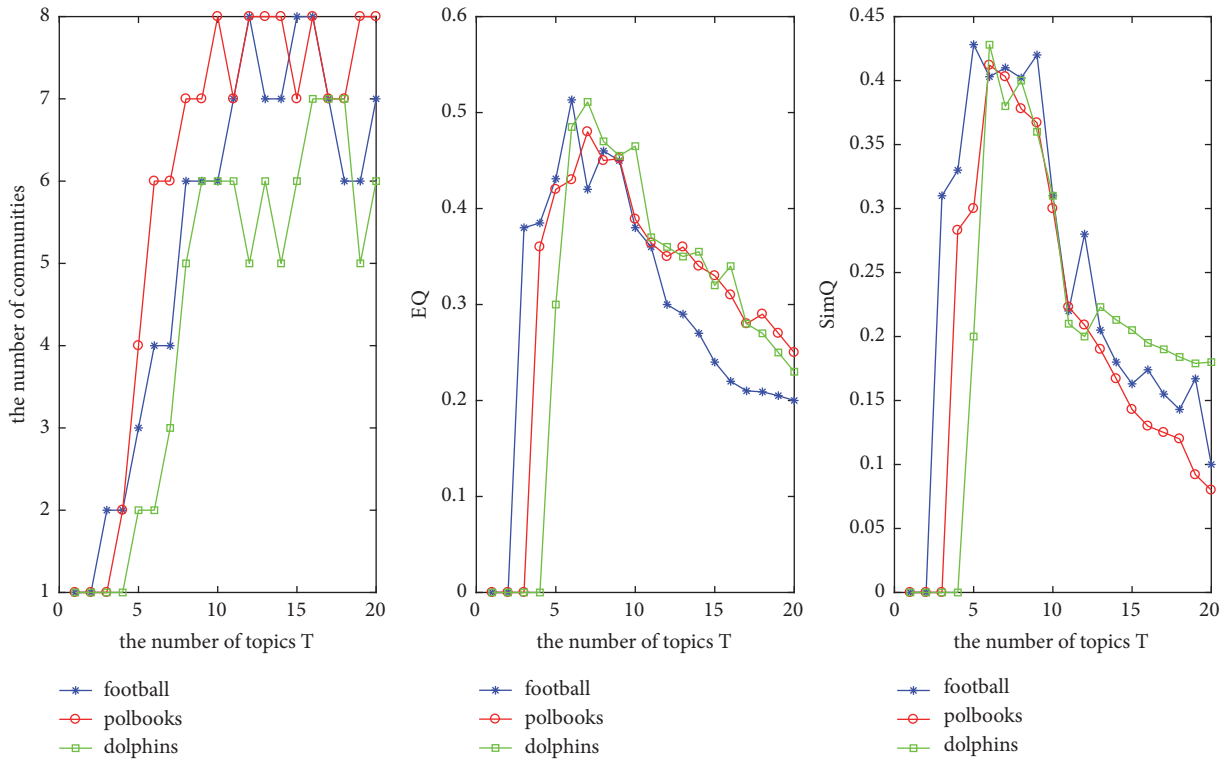
FIGURE 5: The performance of detected communities with T.

$T = 6(football)$ $T = 12(football)$ $T = 18(football)$

$T = 6(polbooks)$ $T = 12(polbooks)$ $T = 18(polbooks)$

$T = 6(dolphins)$ $T = 12(dolphins)$ $T = 18(dolphins)$

(a)   (b)   (c)
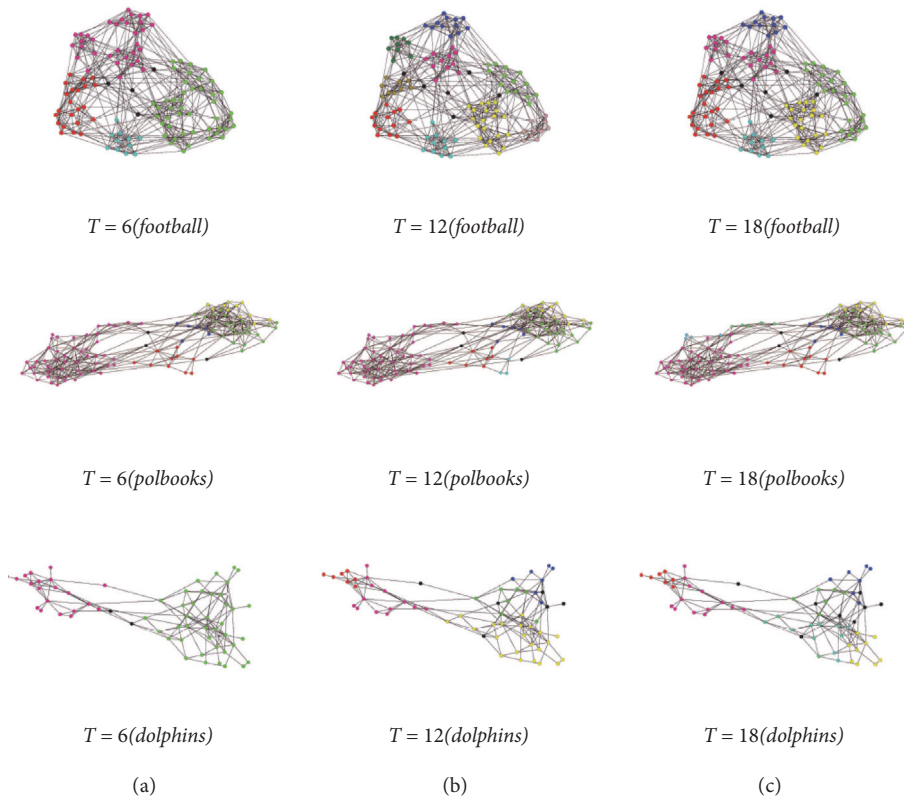
FIGURE 6: The communities for $T = \{6, 12, 18\}$ (the black nodes are overlapping nodes).
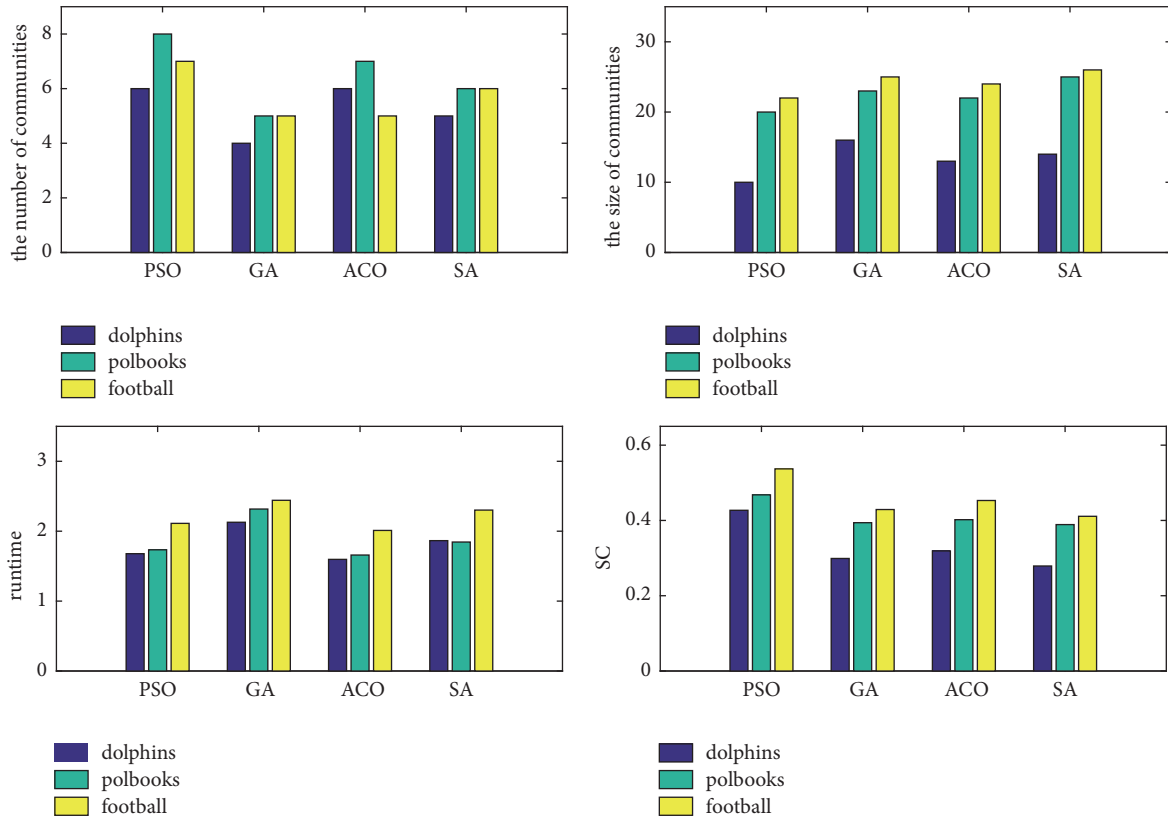
FIGURE 7: The performance of different optimization algorithms.

(a) PSO

(b) GA

(c) ACO

(d) SA
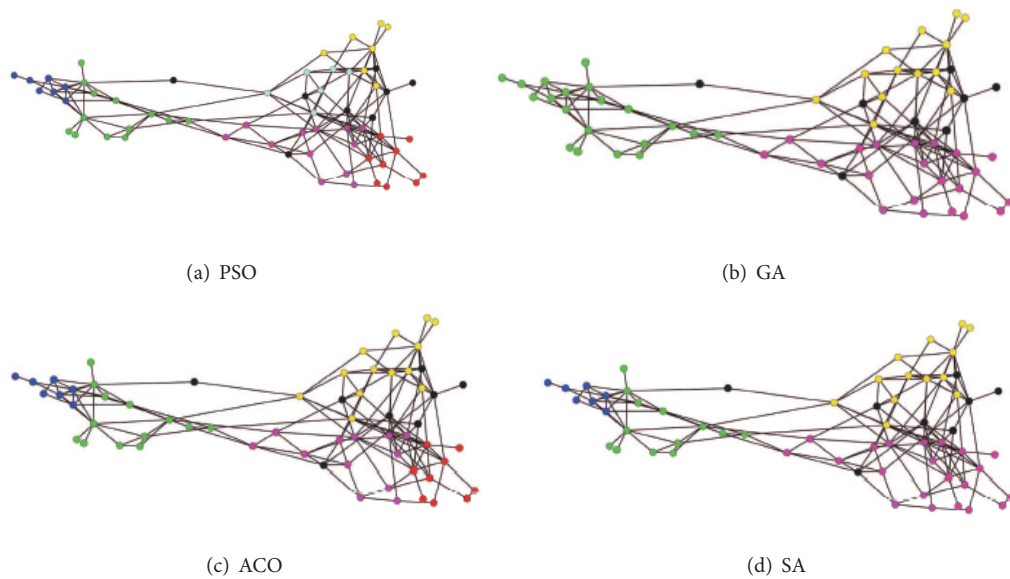
FIGURE 8: The comparison on different optimization algorithms on dolphins (the black nodes are overlapping nodes).

in Figure 7, the detected communities by PSO have a little small size and a bit more community numbers, which is in accordance with the topic distribution. As for runtime, PSO runs a bit slower than ACO but much better than GA and SA. Figure 8 shows four optimization algorithms run on dolphins network, and as similar as Figure 7, PSO works much better than other algorithms on community detection.
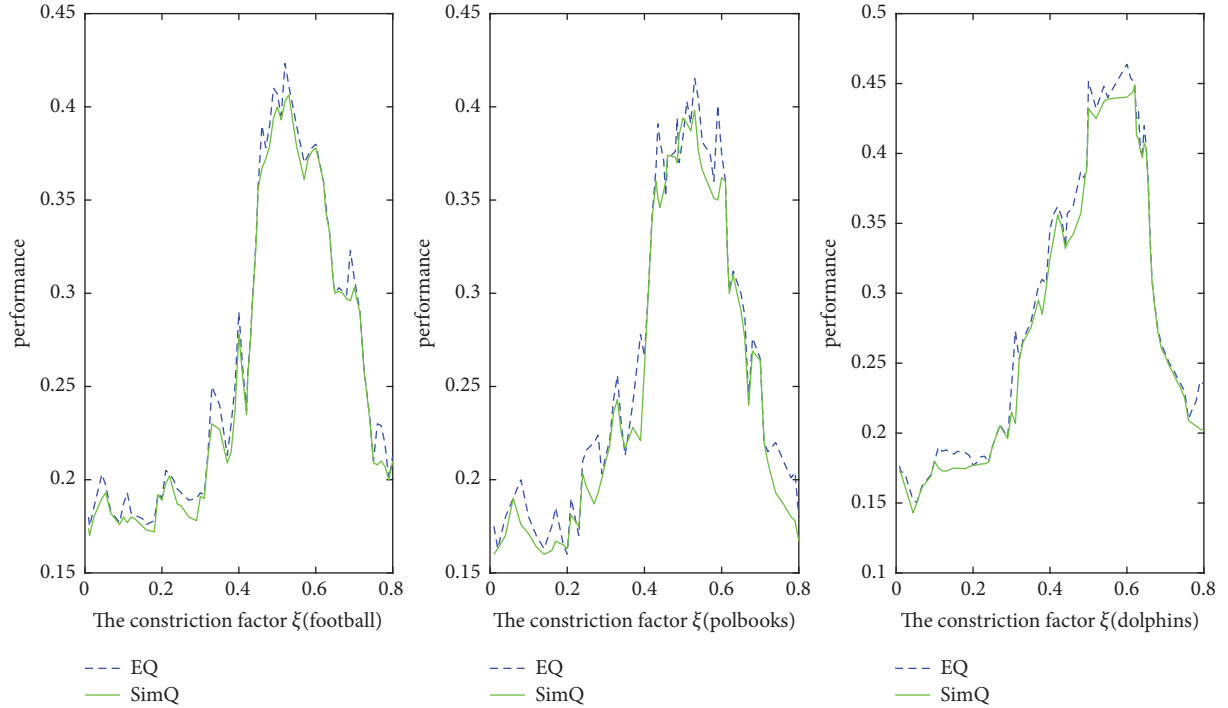
FIGURE 9: The digrams of comparison on the constriction factor with *EQ* and *SimQ*.

TABLE 2: The classical nonsemantic algorithms on *EQ*, *SimQ*, and *SC*.

Algorithms	EQ	SimQ	SC
GN	0.4615	0.3573	0.3873
FN	0.4061	0.3174	0.4012
LFM	0.3255	0.2331	0.3625
COPRA	0.5407	0.4115	0.3902
PSO-LDA	0.5132	0.4258	0.4842

5.3. The Comparison on the Constriction Factor with EQ and SimQ. In this section, we compare *EQ* and *SimQ* over three datasets. The run diagrams, which *EQ* and *SimQ* run in three datasets, are shown in Figure 9. From (16), we put the similar function of information $Sim(d_i, d_j)$ into *SimQ* and $Sim(d_i, d_j) < 1$. So generally, the tendency of *EQ* diagram can be higher than *SimQ*. The maximum value of *EQ* in football dataset is 0.4233 (ξ =0.52) and *SimQ* is 0.4064 (ξ=0.53); and there exists bias when ξ =0.53, and the value of *EQ* is 0.4112 (not the maximum one). There is also bias in polbooks dataset and dolphins dataset, and the maximum value of *EQ* is 0.4154 (ξ =0.54) and *SimQ* is 0.3982 (ξ =0.55) in polbooks dataset while the maximum value of *EQ* is 0.4639 (ξ =0.60) and *SimQ* is 0.4489 (ξ =0.62) in dolphins dataset.

5.4. The Comparison on Community Detection Algorithms. Considering the bias in the semantic community detection, we utilize classical nonsemantic algorithms to illuminate the issue with the football dataset, for example.

We choose GN, FN, LFM, COPRA as nonsemantic classical algorithms, where LFM and COPRA are the overlapping community detection algorithms. The *EQ* and *SimQ* of the algorithms above are covered in Table 2 and the detection of communities is shown in Figure 10 with football dataset.

From the result in Table 2, the *EQ* of nonsemantic classical algorithms work higher than that of PSO-LDA (0.5132), but the *SimQ* works lower than PSO-LDA (0.4258). So it suggests that the nonsemantic classical algorithms make a higher *EQ* in the topological construction detection and a lower *SimQ* in the semantic detection. There is a bias in community detection by nonsemantic classical algorithms compared to semantic algorithms in the way of getting the ideal communities. On the one hand, we verify the performance of these algorithms; on the other hand, we use this experiment to verify the relation above *EQ*, *SimQ*, and *SC*. As for *SC* in Table 2, PSO-LDA performs better in *SimQ* and has high *EQ*, and PSO-LDA is higher than other algorithms in *SC*. This means PSO-LDA performs well in overall search (*EQ* and *SimQ*) and works better than others in local search (*SC*).

5.5. The Comparison on Real Datasets. In this section, we compare real different datasets, including Quantifying Link Semantics-Publication (QLSP) dataset (805 nodes), Academic Social Network (ASN) dataset (extract 2500 nodes) (https://www.aminer.cn/aminernetwork), extracting 10000 nodes and 20000 nodes from DBLP (December 31, 2014) dataset (2839219 nodes) (http://dblp.uni-trier.de/ db/) as DBLP(A) and DBLP(B), and Enron email network (Enron)

(a) GN

(b) FN

(c) LFM

(d) COPRA
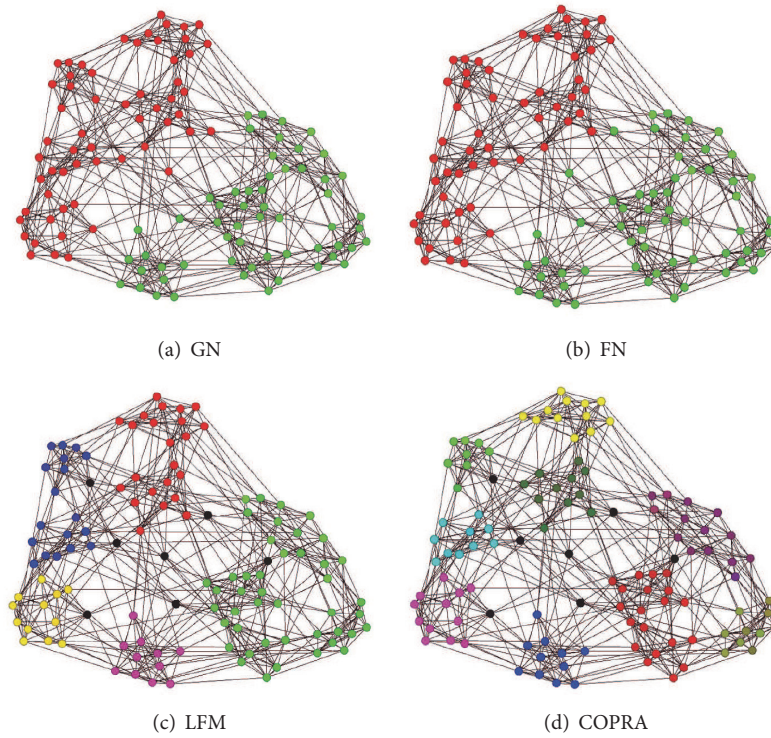
FIGURE 10: The detected communities with nonclassical algorithms on football.
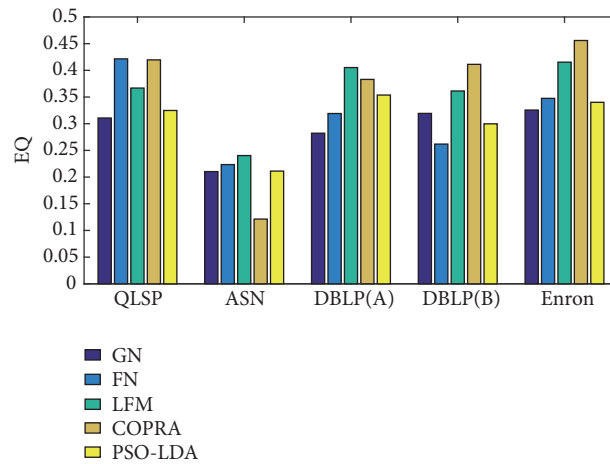
FIGURE 11: The histogram of *EQ* with various classical algorithms.

dataset (extract 25000 nodes) (http://snap.stanford.edu/data/email-Enron.html). The *EQ*, *SimQ*, and *NC*(the number of detected communities) of datasets above detected by various algorithms are reported in Table 3, as the PSO-LDA for $T = 6$. The histogram of *EQ* is shown in Figure 11 and *SimQ* in Figure 12. From Figures 11 and 12, the PSO-LDA model can be more suitable to solve the semantic community detection than the classical nonsemantic algorithms.

6. Conclusion

In this paper, we presented a novel community detection algorithm PSO-LDA that combines the topological construction with semantic information. It can avoid the number and the size of communities. For the Gibbs sampling solving the hidden parameter in the proposed model, the sampling result approaches to the realistic state. The main contribution of this research focuses on how to use different similarity measure to
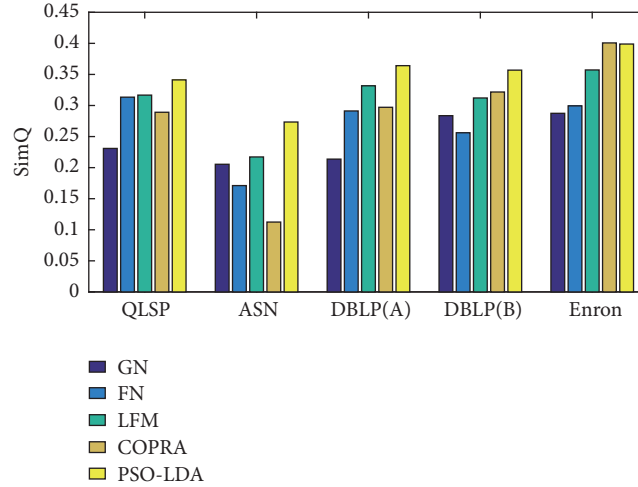
FIGURE 12: The histogram of *SimQ* with various classical algorithms.

TABLE 3: The results of classical nonsemantic algorithms under various datasets.

Algorithms	EQ/SimQ/NC	QLSP	ASN	DBLP(A)	DBLP(B)	Enron
	EQ	0.3107	0.2103	0.2822	0.3193	0.3256
GN	*SimQ*	0.2309	0.2054	0.2137	0.2863	0.2874
	NC	10	35	17	16	27
	EQ	0.4215	0.2234	0.3191	0.2618	0.3475
FN	*SimQ*	0.3134	0.1711	0.2912	0.2561	0.2994
	NC	10	33	19	16	26
	EQ	0.3668	0.2403	0.4052	0.3613	0.4153
LFM	*SimQ*	0.3167	0.2172	0.3317	0.3121	0.3572
	NC	12	29	21	12	30
	EQ	0.4196	0.1213	0.383	0.4112	0.4559
COPRA	*SimQ*	0.2891	0.1124	0.2971	0.3217	0.4007
	NC	13	31	21	13	26
	EQ	0.3248	0.2112	0.3537	0.2998	0.3401
PSO-LDA	*SimQ*	0.3412	0.2734	0.3641	0.3569	0.3989
	NC	14	30	23	15	27

measure similarity between nodes based on topological construction and their semantic information. As for future work, we would apply the model in some fields such as privacy protection and worm containment in semantic social network.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] S. Fortunato and D. Hric, "Community detection in networks: a user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.

[2] Y. Ruan, D. Fuhry, and S. Parthasarathy, "Efficient community detection in large networks using content and links," in *Proceedings of the International Conference on World Wide Web*, pp. 1089–1098, 2013.

[3] U.-U. Narantsatsralt and S. Kang, "Social network community detection using agglomerative spectral clustering," *Complexity*, vol. 2017, Article ID 3719428, 10 pages, 2017.

[4] C.-D. Wang, J.-H. Lai, and P. S. Yu, "NEIWalk: Community discovery in dynamic content-based networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 7, pp. 1734–1748, 2014.

[5] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 70, no. 2, Article ID 066111, 2004.

[6] M. E. Newman, "Fast algorithm for detecting community structure in networks," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, article 066133, 2004.

[7] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 69, no. 2, article 026113, 2004.

[8] G. Palla, I. Dere Nyi, I. S. Farkas, and T. S. Vicsek, "Uncovering the overlapping community structure," *Nature*, vol. 435, no. 7043, pp. 398–406, 2005.

[9] A. Lancichinetti, S. Fortunato, and J. Kertesz, "Detecting the overlapping and hierarchical community structure of complex networks," *New Journal of Physics*, vol. 11, no. 3, pp. 19–44, 2012.

[10] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, pp. 155–168, 2008.

[11] H. A. Deylami and M. Asadpour, "Link prediction in social networks using hierarchical community detection," in *Information and Knowledge Technology*, pp. 1–5, 2015.

[12] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 905–918, 2014.

[13] Y. Neuman, Y. Neuman, and Y. Cohen, "A novel procedure for measuring semantic synergy," *Complexity*, vol. 2017, Article ID 5785617, 8 pages, 2017.

[14] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *Proceedings of the 13th IEEE International Conference on Data Mining (ICDM '13)*, pp. 1151–1156, 2013.

[15] A. Reihanian, M. R. Feizi-Derakhshi, and H. S. Aghdasi, "Overlapping community detection in rating-based social networks through analyzing topics, ratings and links," *Pattern Recognition*, 2018.

[16] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[17] F. Solera, S. Calderara, and R. Cucchiara, "Socially constrained structural learning for groups detection in crowd," *IEEE Computer Society*, 2016.

[18] X. Li, C. Ming, and J. She, "Connection discovery using shared images by gaussian relational topic model," in *Proceedings of the IEEE International Conference on Big Data*, pp. 931–936, 2017.

[19] Y. Xiao, L. Liu, M. Xu, H. Wang, and Y. Liu, "Glda-fp: Gaussian lda model for forward prediction," in *Proceedings of the International Conference on Big Data*, pp. 124–139, 2018.

[20] X. Yu, J. Yang, and Z. Q. Xie, "A semantic overlapping community detection algorithm based on field sampling," *Expert Systems with Applications*, vol. 42, no. 1, pp. 366–375, 2015.

[21] S. Gupta and P. Kumar, "Community detection in heterogenous networks using incremental seed expansion," in *Proceedings of the 2016 International Conference on Data Science and Engineering (ICDSE)*, pp. 1–5, 2017.

[22] Y. Zhao, E. Levina, and J. Zhu, "Consistency of community detection in networks under degree-corrected stochastic block models," *The Annals of Statistics*, vol. 40, no. 4, pp. 2266–2292, 2012.

[23] W. B. Towne, C. P. Rosé, and J. D. Herbsleb, "Measuring similarity similarly: Lda and human perception," *ACM Transactions on Intelligent Systems and Technology*, vol. 8, no. 1, 2016.

[24] S. Cavallari, V. W. Zheng, H. Cai, K. C.-C. Chang, and E. Cambria, "Learning community embedding with community detection and node embedding on graphs," in *Proceedings of the 26th ACM International Conference on Information and Knowledge Management, (CIKM '17)*, pp. 377–386, 2017.

[25] Z. Yin, L. Cao, Q. Gu, and J. Han, "Latent community topic analysis: Integration of community discovery with topic modeling," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 4, pp. 1–21, 2012.

[26] Z. Xia and Z. Bu, "Community detection based on a semantic network," *Knowledge-Based Systems*, vol. 26, pp. 30–39, 2012.

[27] F. Zhao, Y. Zhu, H. Jin, and L. T. Yang, "A personalized hashtag recommendation approach using LDA-based topic model in microblog environment," *Future Generation Computer Systems*, vol. 65, pp. 196–206, 2016.

[28] S. Ahajjam, M. El Haddad, and H. Badir, "A new scalable leader-community detection approach for community detection in social networks," *Social Networks*, vol. 54, pp. 41–49, 2018.

[29] X. Yang and J. Cao, "A Fast and accurate way for API network construction based on semantic similarity and community detection," in *Proceedings of the IFIP International Conference on Network and Parallel Computing*, pp. 75–86, 2017.

[30] C. X. Zhai, "Probabilistic topic models for text data retrieval and analysis," in *Proceedings of the International ACM SIGIR Conference*, pp. 1399–1401, 2017.

[31] G. Heinrich, "Parameter estimation for text analysis," Technical Report, 2008.

[32] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[33] M. Sachan, D. Contractor, T. A. Faruquie, and V. L. Subramaniam, "Using content and interactions for discovering communities in social networks," in *Proceedings of the International Conference on World Wide Web*, pp. 331–340, 2012.

[34] G.-J. Qi, C. C. Aggarwal, and T. Huang, "Community detection with edge content in social media networks," in *Proceedings of the IEEE 28th International Conference on Data Engineering, (ICDE '12)*, pp. 534–545, 2012.

[35] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, 1995.

[36] S. Kianian, M. R. Khayyambashi, and N. Movahhedinia, "Semantic community detection using label propagation algorithm," *Journal of Information Science*, vol. 42, no. 2, pp. 166–178, 2016.

[37] H. Abadlia, N. Smairi, and K. Ghedira, "Particle swarm optimization based on island models," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 49-50, 2017.

[38] X. Wang, D. Jin, X. Cao, L. Yang, and W. Zhang, "Semantic community identification in large attribute networks," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence, (AAAI '16)*, pp. 265–271, 2016.

[39] N. A. Helal, R. M. Ismail, N. L. Badr, and M. G. Mostafa, "An efficient algorithm for community detection in attributed social networks," in *Proceedings of the International Conference on Informatics and Systems*, pp. 180–184, 2016.

[40] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706–1712, 2009.

[41] A. Clauset, "Finding local community structure in networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 72, no. 2, Article ID 026132, 2005.

Research Article
Variational Approach for Learning Community Structures

Jun Jin Choong [iD],[1] Xin Liu,[2] and Tsuyoshi Murata[1]

[1]*Department of Computer Science, Tokyo Institute of Technology, Tokyo, Japan*
[2]*National Institute of Advanced Industrial Science and Technology, Tokyo, Japan*

Correspondence should be addressed to Jun Jin Choong; choong.junjin@gmail.com

Discovering and modeling *community structure* exist to be a fundamentally challenging task. In domains such as biology, chemistry, and physics, researchers often rely on community detection algorithms to uncover community structures from complex systems yet no unified definition of community structure exists. Furthermore, existing models tend to be oversimplified leading to a neglect of richer information such as nodal features. Coupled with the surge of user generated information on social networks, a demand for newer techniques beyond traditional approaches is inevitable. Deep learning techniques such as network representation learning have shown tremendous promise. More specifically, supervised and semisupervised learning tasks such as link prediction and node classification have achieved remarkable results. However, unsupervised learning tasks such as community detection remain widely unexplored. In this paper, a novel deep generative model for community detection is proposed. Extensive experiments show that the proposed model, empowered with Bayesian deep learning, can provide insights in terms of uncertainty and exploit nonlinearities which result in better performance in comparison to state-of-the-art community detection methods. Additionally, unlike traditional methods, the proposed model is community structure definition agnostic. Leveraging on low-dimensional embeddings of both network topology and feature similarity, it automatically learns the best model configuration for describing similarities in a community.

1. Introduction

Real-world complex systems are often projected into networks to observe complex patterns. Entities in a complex system can be represented as nodes (vertices) and their interactions represented as an edge (link). For instance, social interactions between people can be represented in the form of a social network. Publications by authors and their respective publication venues can be represented with a bipartite citation network. The flexibility of networks and its vast literature on graph theory make network science very appealing to researchers. Although networks are merely represented in forms of nodes and edges, a large complex system could easily scale from hundreds to millions of nodes and edges. This poses a very challenging task in machine learning, especially tasks such as graph clustering or more commonly known as community detection [1] in the literature of network science. Given a network (graph) with its node content and structural (link) information, community detection aims to partition the nodes in the network into a number of disjoint groups. These partitions can be formulated depending on the given definition. For example, in modularity maximization [2], each partition is compared against a null model (random network). A partition is classified as good when the modularity score is greater than partitioning a random network. On the other hand, statistical methods such as the Stochastic Blockmodel (SBM) introduced Bayesian treatment of uncertainty when partitioning the network. Nodes with similar statistical similarity have higher probability to cluster together regardless of the cluster's density [3]. This is known as stochastic equivalence. In general, a universal definition of community structure does not exist. Nevertheless, the objective remains the same, i.e., to find a group of nodes that shares some form of similarity between one another. In this paper, such similarity is defined as latent similarity; the similarity measure is not predefined. Quantifying such similarity is arguably subjective and difficult especially when a given network can be feature-rich or structure-only; there is no one-size-fits-all solution for community detection (i.e., the *no free lunch theorem*). Therefore, it is essential

that algorithms capture both higher-order information and structural information. To this end, we look at network representation learning [4, 5] as a potential solution.

In machine learning, representation learning [6] has been successfully applied to various fields such as natural language processing and computer vision. Notably, successes of deep learning have surpassed human accuracy with ease [7]. However, these successes are difficult to be explained. More precisely, it is difficult to explain "why" deep learning model performs so well. In an attempt to solve this problem, researchers bridged the understanding gap by introducing probabilistic deep models (also known as Bayesian Deep Learning) [8]. Using fundamental building blocks from a probabilistic perspective, assumptions are given in forms of noninformative priors and the model is forced to correct these assumptions while learning. Consequently, the models become less ambiguous than a typical deep learning model which is commonly known to be a black-box.

Leveraging on recent advances in representation learning, network representation learning aims at a similar objective, but from a network perspective. Given a network, the objective is to find a latent representation that generalizes for various machine learning tasks such as classification, link prediction, and clustering of nodes. Generally, a common choice for finding community structure in networks often involves a two-step approach. First, the network is embedded into a latent space (i.e., Euclidean space). Next, a general clustering algorithm such as Spectral Clustering [9] or k-means is applied to the learned embedding. For instance, Tian *et al.* proposed a network representation [10] learning model to learn a nonlinear mapping of the original network using a Stacked Autoencoder by showing that spectral clustering and Autoencoders have the same optimization objectives. Yang *et al.* considered a Stacked Autoencoder as a modularity optimization problem and further introduced a semisupervised approach through *must-pair* nodes for increased performance [11]. Assignment of communities is then obtained through k-means clustering from the latent representation that exhibits the highest modularity score. Inspired from Denoising Autoencoders [12], Wang *et al.* proposed Marginalized Graph Autoencoder for Graph Clustering (MGAE) [13] that artificially corrupts the feature matrix to increase the number of training data and provides a close-form solution for optimization. Spectral Clustering is then applied to the learned latent representation. Clearly, these methods all employ a two-step approach which is unsuitable for studying network generation or graph modeling [14].

Instead of a costly two-step approach and ignoring uncertainty in the modeling process, the problem can be solved from a Bayesian point of view, by encoding our latent beliefs and assumptions as probabilistic graphical models. Specifically, one can assume that nodes and edges are modeled from a mixture model such as the Gaussian Mixture Model (GMM). This effectively couples the learning of cluster assignment with respect to its network representation into a joint probability distribution. Additionally, it helps to capture network properties exhibited by common networks which consequently helps in better understanding of real-world networks.

Concretely, this paper proposes an extension to Variational Graph Autoencoder (VGAE) [15]. Originally, VGAE projects graph convolutions into a Univariate Gaussian latent space and have only been considered for semisupervised task such as link prediction and graph classification. The proposed model, VGAECD, relaxes this notion by introducing a Mixture of Gaussian. This is desirable as we would like to capture higher-order patterns from community structures and model its generative process. It is worth noting that similar approaches have been applied to VAE in domains such as image recognition [16]. However, these approaches are not readily applicable for networks, especially in a community detection problem.

To summarize, this paper explores the idea of learning network representations using Bayesian treatment. We extend VGAE to include clustering-aware capability specifically targeting a community detection task. The contribution of this paper is summarized as follows:

(i) This paper proposes a novel generative model for community detection which is agnostic to the necessity of a predefined community structure definition. Through the process of automatic model selection, nodes are assigned a community based on the criterion that best reduces the loss function.

(ii) The proposed model inherits the benefits of Variational Autoencoder Framework. The advantages are threefold: (1) it provides a variational lower bound which is guaranteed to converge to a local minimum, (2) the lower bound is scalable, and (3) the model is generative, allowing generation of synthetic networks.

(iii) The proposed model outperforms the state-of-the-art models in community detection without requiring additional priors (unlike the Degree-Corrected SBM).

2. Problem Definition

A network pertaining to nodes, edges, and node features can be formally defined as $G = (V, E, X)$, where $V = \{v_i, \ldots, v_N\}$ consists of a set of nodes $|V| = N$, $E = \{e_{ij}\}$ is a set of edges, and $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ is the set of node features. Each $\mathbf{x}_i \in \mathbb{R}^d$ defines a vector of real-values associated with node v_i. From an Autoencoder's perspective, the inputs are given in terms of structural information $A \in \mathbb{R}^{N \times N}$, and node features $X \in \mathbb{R}^{N \times d}$, where A denotes the adjacency matrix of G, and the node features are content information provided in forms of vector representation. In this work, we consider the undirected and unweighted network G, such that $A_{ij} = 1$ if $e_{ij} \in E$ and otherwise it is equal to 0.

Given the network G, the objective of community detection or graph clustering is to partition the nodes in G into K disjoint groups $\{c_1, c_2, \ldots, c_K\}$, such that nodes grouped within the same cluster are close to each other while nodes in different clusters are distant in terms of network structure. Vertices grouped within the same cluster are more likely to have similarities in node features.

Additionally, we consider the definition of a generative model. The discriminative model, $p(\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{A})$, infers the

model parameters $\boldsymbol{\theta}$ from the observed network G. Subsequently, a network G' can be generated from the same set of parameters. Concretely, $p(\mathbf{A} \mid \boldsymbol{\theta}) = G'$. Under the model selection criterion, the model is said to be good when $G' \cong G$ and satisfies the condition of having community structures; i.e., G' is not an Erdős–Rényi network. By definition, generative models can be considered as an ensemble learning model.

3. Related Work

Recent work in community detection can be broadly categorized into two types of models, namely, discriminative and generative models. The former includes a class of methods that infers communities given an observed network and, optionally, node features. Meanwhile, the latter considers the reconstruction of network while exploring plausible models that explain the observed phenomenon.

3.1. Discriminative Methods and Models. Predominantly, modularity maximization [2, 17] has been considered as the most successful method for detecting communities. However, it suffers from a resolution limit problem [18] and is known to exhibit degeneracies [19]. In terms of speed, label propagation [20] is capable of detecting communities in large-scale networks near linear time, though the solutions are usually nonunique. Additionally, other approaches such as Walk-Trap [21], Infomap [22], Louvain [23], and their empirical competitiveness are subjected to trade-off between accuracy and scalability [24]. Representation learning methods such as GraRep [25] and CFOND [26] consider the completion of their adjacency matrix and can be generally considered as matrix factorization problem. Meanwhile, others like DeepWalk [27] and node2vec [28] consider representation of each node via a biased random walk. It assumes that neighboring nodes share similarities from the pivot node. Hence, when nodes are clustered together, they tend to co-occur on short random walks over the network.

Besides standard linear methods mentioned previously, recent advances in deep learning revisited Autoencoders for networks. Particularly, GraphEncoder proposed by Tian *et al.* shows that optimizing the objective function of Autoencoder is similar to finding a solution for Spectral Clustering [10]. Leveraging on deep learning's nonlinearity and recent advances in Convolutional Neural Networks, [29, 30] proposed the Graph Neural Network (GNN) and its generalization, the Graph Convolutional Neural Network (GCN) [29]. Defferrard *et al.* first cast the problem by projecting graph convolutions into spectral space, and convolving within this space.

3.2. Generative Methods and Models. Generative models can be further subdivided into algorithmic and statistical types. Examples of algorithmic models include the Kronecker Graphs [31], NetSim [32], and Block Two-Level Erdős-Rényi (BTER) model [33]. On the other hand, statistical methods attempt to approximate the true distribution via statistical inferencing or through statistical models (i.e., benchmark graphs such as GN [34], LFR [35], and mLFR [36, 37]).

A widely known generative model for capturing networks with group structure is the Stochastic Blockmodel (SBM) or also known as the planted partition model. First explored by Snijders and Nowicki [38] two decades ago, the key idea behind SBM is stochastic equivalence. The probability that two nodes i and j are connected depends exclusively on their community memberships: two nodes within a community sharing the same stochasticity. However, the vanilla SBM exhibits a problem where high degree nodes are clustered into a community of their own. Karrer and Newman proposed the Degree Corrected (D.C.) SBM [39] which introduces a normalizing prior. Extensions to SBM include the Mixed Membership SBM (MMSBM) [40] for identifying mix community participation and bipartite SBM (biSBM) [41] for finding communities in bipartite networks. Today, SBM is well explored and its limitations has been widely studied [42, 43]. However, SBM is not a network representation learning model. Instead, SBM learns the latent variables $\boldsymbol{\Pi}$ and $\mathbf{Z}$ which describe the probabilities of cluster connectivity and cluster assignment, respectively, of a particular node which differs from common representation learning method.

Contrary to SBM, typically Autoencoders consists of two nongenerative steps (encoder and decoder). Consequently, the learned representation cannot be generalized for generation of networks. To alleviate this problem, most recent approaches consider generative models for representation learning such as Generative Adversarial Networks (GAN) or Variational Autoencoder (VAE). For graphs, Kipf and Welling [15] introduced a variant of VAE for link prediction tasks in graphs and for GAN, and Pan *et al.* [44] recently introduced adversarially regularized graph autoencoder (ARGA). In this work, we only consider the framework of VAE. We discuss this in Section 4.1.

4. Methodology

4.1. Variational Graph Autoencoder. Variational Graph Autoencoder (VGAE) [15] extends the problem of learning network embedding to a generative perspective by leveraging on the Variational Autoencoder (VAE) framework [45]. Consider a given network G with structural information $\mathbf{A}$ and node features $\mathbf{X}$; the inference model of VGAE parameterized by a two-layer GCN is defined as

$$q\left(\mathbf{Z} \mid \mathbf{X}, \mathbf{A}\right) = \prod_{i=1}^{N} q\left(\mathbf{z}_i \mid \mathbf{X}, \mathbf{A}\right) \qquad (1)$$

$$q\left(\mathbf{z}_i \mid \mathbf{X}, \mathbf{A}\right) = \mathcal{N}\left(\mathbf{z}_i \mid \boldsymbol{\mu}_i, \operatorname{diag}\left(\boldsymbol{\sigma}_i^2\right)\right). \qquad (2)$$

Here, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ denote the mean and standard deviation vectors for node i which is obtained from a GCN layer, $\boldsymbol{\mu} = \operatorname{GCN}_{\boldsymbol{\mu}}(\mathbf{X}, \mathbf{A})$ and $\log \boldsymbol{\sigma} = \operatorname{GCN}_{\boldsymbol{\sigma}}(\mathbf{X}, \mathbf{A})$. The two-layer GCN is then defined as

$$\operatorname{GCN}\left(\mathbf{X}, \mathbf{A}\right) = \widehat{\mathbf{A}}\tau\left(\widehat{\mathbf{A}}\mathbf{X}\mathbf{W}_0\right)\mathbf{W}_1, \qquad (3)$$

with $\mathbf{W}_0$ and $\mathbf{W}_1$ representing the weight matrices for the first layer and second layer, respectively. $\mathbf{W}_0$ is shared between $\operatorname{GCN}_{\boldsymbol{\mu}}(\mathbf{X}, \mathbf{A})$ and $\operatorname{GCN}_{\boldsymbol{\sigma}}(\mathbf{X}, \mathbf{A})$. $\tau(\cdot)$ is the nonlinear function

such as $\text{ReLU}(\cdot) = \max(0, \cdot)$ or $\text{sigmoid}(t) = 1/(1 + e^{-t})$. $\widehat{\mathbf{A}} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ denotes the symmetric normalized adjacency matrix. The generative model is simply the inner product between the latent variables:

$$p(\mathbf{A} \mid \mathbf{Z}) = \prod_{i=1}^{N}\prod_{j=1}^{N} p\left(A_{ij} = 1 \mid \mathbf{z}_i \mathbf{z}_j\right) = \tau\left(\mathbf{z}_i^{\top}\mathbf{z}_j\right). \quad (4)$$

In accordance to the VAE framework, both models can be tied together and optimized by maximizing the variational lower bound $\mathscr{L}(\cdot)$:

$$\log p_{\boldsymbol{\theta}}(\mathbf{X}) \geq \mathscr{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{X})$$
$$= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{Z}|\mathbf{X},\mathbf{A})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{A} \mid \mathbf{Z})\right] \quad (5)$$
$$- D_{KL}\left[q_{\boldsymbol{\phi}}(\mathbf{Z} \mid \mathbf{X}) \parallel p_{\boldsymbol{\theta}}(\mathbf{Z})\right].$$

$D_{KL}[q_{\boldsymbol{\phi}}(\cdot) \parallel p_{\boldsymbol{\theta}}(\cdot)]$ defines the Kullback-Leibler (KL) divergence between $q_{\boldsymbol{\phi}}(\cdot)$ and $p_{\boldsymbol{\theta}}(\cdot)$. The lower bound can be maximized with respect to the variational parameters $(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathbf{W}_i$ via stochastic gradient descent, performed with a full-batch size. Here, the prior is defined as $p_{\boldsymbol{\theta}}(\mathbf{Z}) = \prod_{i=1}^{N}\mathcal{N}(\mathbf{z}_i \mid 0, \mathbf{I})$, which is the isotropic Gaussian distribution, whose gradients can backpropagate via a *reparametrization trick* [45].

In the absence of node features, $\mathbf{X}$ becomes the identity matrix. This relaxation allows the reconstruction of a structure-only network. When provided with node features, the accuracy of VGAE link prediction improves [15].

4.2. Variational Graph Autoencoder for Community Detection (VGAECD). A major drawback in VGAE's approach is its restriction of nodes to be projected in a Univariate Gaussian space. This restriction suggests that all generated nodes come from a single clustering space. More specifically, dissimilar nodes tend to stay away from the Gaussian mean (centroid) [15]. On the contrary, the mean of the Gaussian should be a better representative of each respective community such that nodes which are similar should stay closer to their represented mean. Thus, nodes that are well represented by the mean representation hold equivalence in similarity. In this scenario, we can consider this as a relaxation of SBM which requires nodes in the same block to uphold stochastic equivalence.

Utilizing this fact, we consider the unsupervised learning problem of community detection while adhering to the VGAE framework. Suppose that each node originating from a particular community is similar in some way; we can encode their similarity into the node's representation vector $\mathbf{z}$ which is better described by the mixture's mean. The generative process then follows:

(i) For communities $C = \{c_1, \ldots, c_K\}$

    (a) Obtain a sample $c \sim \text{Cat}(\boldsymbol{\pi})$

    (b) where K is the number of clusters hyperparameters and π_k is the prior probability for cluster k, $\boldsymbol{\pi} \in \mathbb{R}_{+}^{K}$, $\sum_{k=1}^{K}\pi_k = 1$. $\text{Cat}(\boldsymbol{\pi})$ is the categorical distribution parameterized by $\boldsymbol{\pi}$.

(ii) For nodes $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$,

    (a) Obtain a latent vector $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\sigma}_c^2 \mathbf{I})$

    (b) where $\boldsymbol{\mu}_c$ and $\boldsymbol{\sigma}_c^2$ are the mean and variance of the multivariate Gaussian distribution corresponding to cluster c.

(iii) Obtain a sample $\mathbf{a}$ by

    (a) computing the expectation $\boldsymbol{\mu}_x = f(\mathbf{z}; \boldsymbol{\theta})$

    (b) sample $\mathbf{a} \sim \text{Bern}(\boldsymbol{\mu}_x)$

The function $f(\mathbf{z}; \boldsymbol{\theta})$ is optionally a nonlinear function whose input is $\mathbf{z}$ and is parameterized by $\boldsymbol{\theta}$. Particularly, we use the $\tau(\mathbf{z}_i^{\top}\mathbf{z}_j)$ inner product decoder. $\text{Bern}(\cdot)$ denotes the multivariate Bernoulli distribution parameterized by the latent vector $\boldsymbol{\mu}_x$. Then, the joint probability $p(\mathbf{a}, \mathbf{z}, c)$ can be factorized as

$$p(\mathbf{a}, \mathbf{z}, c) = p(\mathbf{a} \mid \mathbf{z})p(\mathbf{z} \mid c)p(c), \quad (6)$$

with $\mathbf{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$. Since $\mathbf{a}$ and c are independently conditioned on $\mathbf{z}$, the factorized probabilities can be defined as

$$p(c) = \text{Cat}(c \mid \boldsymbol{\pi}) \quad (7)$$
$$p(\mathbf{z} \mid c) = \mathcal{N}\left(\mathbf{z} \mid \boldsymbol{\mu}_c, \boldsymbol{\sigma}_c^2 \mathbf{I}\right) \quad (8)$$
$$p(\mathbf{a} \mid \mathbf{z}) = \text{Bern}(\mathbf{a} \mid \boldsymbol{\mu}_x) \quad (9)$$

For brevity, $p_{\boldsymbol{\theta}}(\cdot) = p(\cdot)$ and $q_{\boldsymbol{\phi}}(\cdot) = q(\cdot)$, $\mathscr{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}) = \mathscr{L}_{\text{ELBO}}(\mathbf{x})$; we can rewrite the lower bound in (5) to include the new terms:

$$\log p(\mathbf{x}) \geq \mathscr{L}_{\text{ELBO}}(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z},c|\mathbf{x},\mathbf{a})}\left[\log\frac{p(\mathbf{a}, \mathbf{z}, c)}{q(\mathbf{z}, c \mid \mathbf{x}, \mathbf{a})}\right], \quad (10)$$

where $q(\mathbf{z}, c \mid \mathbf{x}, \mathbf{a})$ is the variational posterior which approximates the true posterior $p(\mathbf{z}, c \mid \mathbf{x}, \mathbf{a})$. Under the mean-field assumption, the approximate distribution can be factorized as

$$q(\mathbf{z}, c \mid \mathbf{x}, \mathbf{a}) = q(\mathbf{z} \mid \mathbf{x}, \mathbf{a})q(c \mid \mathbf{x}, \mathbf{a}). \quad (11)$$

Substituting (6) and (11) into (10), $\mathscr{L}_{\text{ELBO}}(\mathbf{x})$ can be rewritten as

$$\mathscr{L}_{\text{ELBO}}(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z},c|\mathbf{x},\mathbf{a})}\left[\log\frac{p(\mathbf{a}, \mathbf{z}, c)}{q(\mathbf{z}, c \mid \mathbf{x}, \mathbf{a})}\right]$$
$$= \mathbb{E}_{q(\mathbf{z},c|\mathbf{x},\mathbf{a})}\left[\log p(\mathbf{a}, \mathbf{z}, c) - \log q(\mathbf{z}, c \mid \mathbf{x}, \mathbf{a})\right] \quad (12)$$
$$= \mathbb{E}_{q(\mathbf{z},c|\mathbf{x},\mathbf{a})}\left[\log p(\mathbf{a} \mid \mathbf{z}) + \log p(\mathbf{z} \mid c) + \log p(c)\right.$$
$$\left. - \log q(\mathbf{z} \mid \mathbf{x}, \mathbf{a}) - \log q(c \mid \mathbf{x}, \mathbf{a})\right].$$

The inference model $q(\mathbf{z} \mid \mathbf{x}, \mathbf{a})$ is then modeled using a two-layer GCN as follows:

$$q(\mathbf{z} \mid \mathbf{x}, \mathbf{a}) = \mathcal{N}\left(\mathbf{z}; \text{GCN}_{\boldsymbol{\mu}}(\mathbf{x}, \mathbf{a}), \text{GCN}_{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{a})\mathbf{I}\right)$$
$$= \mathcal{N}\left(\mathbf{z}; \widetilde{\boldsymbol{\mu}}, \log\widetilde{\boldsymbol{\sigma}}\mathbf{I}\right). \quad (13)$$

Similar to VGAE, the first layer's weight matrix $\mathbf{W}_0$ is shared between $\tilde{\boldsymbol{\mu}}$ and $\log \tilde{\boldsymbol{\sigma}}$. Substituting the terms, $\mathscr{L}_{\text{ELBO}}(\mathbf{x})$ can be further rewritten as

$$
\begin{aligned}
\mathscr{L}_{\text{ELBO}}(\mathbf{x}) = {} & \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{N}\mathbf{x}_i \log \boldsymbol{\mu}_i^{(l)} \\
& + \left(1 - \mathbf{x}_i\right)\log\left(1 - \boldsymbol{\mu}_i^{(l)}\right) \\
& - \frac{1}{2}\sum_{c=1}^{K}\gamma_c\left(\log\boldsymbol{\sigma}_c^2 + \frac{\tilde{\boldsymbol{\sigma}}^2}{\boldsymbol{\sigma}_c} + \frac{\left(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}_c\right)^2}{\boldsymbol{\sigma}_c^2}\right) \\
& + \sum_{c=1}^{K}\gamma_c \log\frac{\pi_c}{\gamma_c} + \frac{1}{2}\left(1 + \log\tilde{\boldsymbol{\sigma}}_c^2\right),
\end{aligned}
\tag{14}
$$

with L being the total number of samples through sampled using the Monte Carlo Stochastic Gradient Variational Bayes (SGVB) estimator [45]. $\mathbf{x}_i$ is the vector of node i, K is the number of clusters with π_c denoting the prior probability of cluster c, and γ_c denotes $q(c \mid \mathbf{x}, \mathbf{a})$ for brevity. $\boldsymbol{\mu}_x^{(l)}$ is computed as

$$
\boldsymbol{\mu}_x^{(l)} = \tau\left(\mathbf{z}_i^{\top}\mathbf{z}_j\right),
\tag{15}
$$

where $\mathbf{z}^{(l)}$ is the lth sample from $q(\mathbf{z} \mid \mathbf{x}, \mathbf{a})$ as written in (13). To allow gradient backpropagation through the stochastic layer, the *reparameterization trick* is used; then $\mathbf{z}^{(l)}$ can be obtained via

$$
\mathbf{z}^{(l)} = \tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{\sigma}} \circ \boldsymbol{\epsilon}^{(l)}.
\tag{16}
$$

Then, according to [45], $\boldsymbol{\epsilon}^{(l)} \sim \mathcal{N}(0, \mathbf{I})$; $\circ$ is the Hadamard product operator. $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\sigma}}$ are obtained through GCN$(\cdot)$.

If we consider regrouping $\mathscr{L}_{\text{ELBO}}(\mathbf{x})$ with like-terms, (12) can be rewritten as

$$
\begin{aligned}
\mathscr{L}_{\text{ELBO}}(\mathbf{x}) = {} & \mathbb{E}_{q(\mathbf{z},c\mid\mathbf{x},\mathbf{a})}\left[\log\frac{p(\mathbf{a},\mathbf{z},c)}{q(\mathbf{z},c \mid \mathbf{x},\mathbf{a})}\right] \\
= {} & \int_{\mathbf{z}}\sum_c q(\mathbf{z} \mid \mathbf{x},\mathbf{a})\, q(c \mid \mathbf{x},\mathbf{a}) \\
& \cdot\left[\log\frac{p(\mathbf{x},\mathbf{a} \mid \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})} + \log\frac{p(c \mid \mathbf{z})}{q(c \mid \mathbf{x})}\right]d\mathbf{z} \\
= {} & \int_{\mathbf{z}} q(\mathbf{z} \mid \mathbf{x},\mathbf{a})\log\frac{p(\mathbf{x},\mathbf{a} \mid \mathbf{z})\, p(\mathbf{z})}{q(\mathbf{z} \mid \mathbf{x},\mathbf{a})}d\mathbf{z} \\
& - \int_{\mathbf{z}} q(\mathbf{z} \mid \mathbf{x},\mathbf{a})\, D_{KL}\left[q(c \mid \mathbf{x}) \parallel p(c \mid \mathbf{z})\right]d\mathbf{z}.
\end{aligned}
\tag{17}
$$

The first term in (17) has no dependency on c and from the definition of KL divergence, it is nonnegative. Therefore, $\mathscr{L}_{\text{ELBO}}(\mathbf{x})$ is maximized when $D_{KL}[q(c \mid \mathbf{x}) \parallel p(c \mid \mathbf{z})] \equiv 0$. From that, we follow [16], by defining $q(c \mid \mathbf{x}, \mathbf{a})$ as

$$
q(c \mid \mathbf{x},\mathbf{a}) = p(c \mid \mathbf{z}) \equiv \frac{p(c)\, p(\mathbf{z} \mid c)}{\sum_{c'=1}^{K} p(c')\, p(\mathbf{z} \mid c')}
\tag{18}
$$

From (18) the information loss induced by the mean-field approximation can be mitigated by forcing its dependency on the posterior $p(c \mid \mathbf{z})$ and noninformative prior $p(c)$. The complete VGAECD algorithm can be found in Algorithm 1 and Figure 1 illustrates the conceptual idea of VGAECD.

5. Experiments

Community detection algorithms are often evaluated against two kinds of networks: synthetic and empirical datasets. These are discussed in detail in the following subsections.

5.1. Synthetic Datasets. Two synthetic networks are used in our evaluation. We consider two most common benchmark graphs used for benchmarking community detection algorithm. Namely, we used the Girvan-Newman (GN) benchmark graph [1, 34, 46] and the LFR benchmark graph [35]. The GN benchmark graph is a variant of the planted l-partition. In our experiment, we vary the z_{out} value from a range of $\{1,\ldots,8\}$. Each node has an average degree of $k = 16$, with 32 nodes in each community (a total of 128 nodes) and 4 communities in total.

The LFR benchmark graph is an extension of the GN benchmark graph. It is considered to be more realistic than the GN benchmark graph. It introduces a skewed degree distribution and accounts for network heterogeneity, resulting in communities that are generated in different sizes. The LFR benchmark graph is generated using default parameters as suggested by Lancichinetti *et al.* [35]. These parameters are number of nodes ($N = 1000$), average degree ($k = 15$), and minimum ($c_{min} = 30$) and maximum ($c_{max} = 50$) number of nodes per community. The generation follows the *scale-free* parameters settings of exponents $\tau_1 = -2$ and $\tau_2 = -1$, respectively. On average, between 20 and 30 communities are generated.

5.2. Empirical Datasets. The empirical datasets are divided into two kinds: networks with features and without features. The datasets are as follows:

(i) **Karate**: a social network represents friendship among 34 members of a karate club at a US University [47].

(ii) **PolBlogs**: a network of political blogs assembled by Adamic and Glance [48]. The nodes are blogs and web links between them are represented by their edge. These blogs have known political leanings and were labelled by hand by Adamic and Glance.

(iii) **Cora**: a citation network with 2,708 nodes and 5,429 edges. Each node corresponds to a document and the edges are citation links [49].

(iv) **PubMed**: A network consisting of 19,717 scientific publications from PubMed database pertaining to diabetes was classified into one of three classes ("Diabetes Mellitus, Experimental", "Diabetes Mellitus Type 1", "Diabetes Mellitus Type 2"). The citation network consists of 44,338 links. Each publication in the dataset is described by a TF-IDF weighted word vector from a dictionary which consists of 500 unique words.
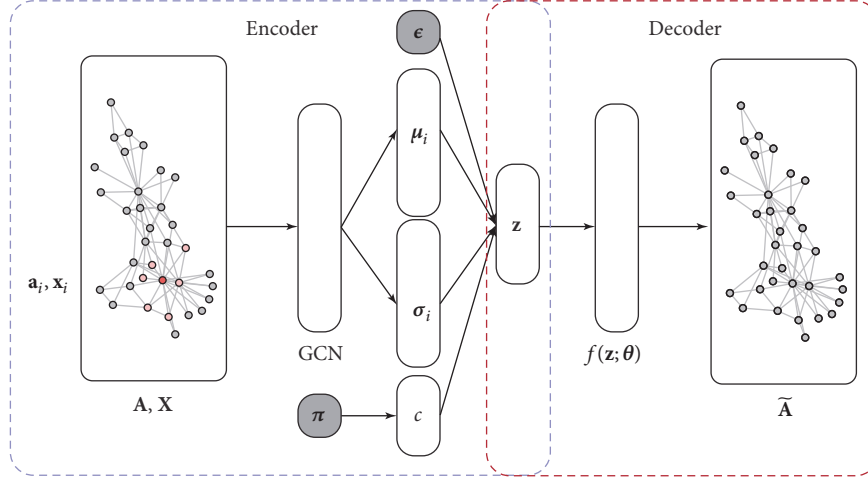
FIGURE 1: Conceptual illustration of Variational Graph Autoencoder Framework for Community Detection (VGAECD). In the encoding phase, VGAECD first convolves on the network, learning structural and nodal features in the process. These pieces of information are then mapped into a latent representation, $\boldsymbol{\mu}_c|_i$ and $\boldsymbol{\sigma}_c|_i$ which are parameters to Mixture of Gaussian Model. Subsequently, we can then sample to obtain a latent representation for each node $\mathbf{z}$. Finally, $\widetilde{\mathbf{A}}$ can be reconstructed using a decoding function, $f(\cdot)$. The loss is calculated and backpropagated to the latent variables.

Input: Features $\mathbf{X}$, Adjacency matrix $\mathbf{A}$,
    Hyperparameters: learning rate ϵ, epochs L, size of layer 1 and 2.
Output: Community Assignment Probability $\boldsymbol{\gamma}$ and Reconstructed Adjacency matrix $\widetilde{\mathbf{A}}$
    $\boldsymbol{\pi} \sim \mathcal{U}(0, 1)$
    for $l = 1, ..., L$ **do**
        for $i = 1, ..., N$ **do**
            $\boldsymbol{\mu}_i = \text{GCN}_{\boldsymbol{\mu}}(\mathbf{x}_i, \mathbf{a}_i)$
            $\boldsymbol{\sigma}_i = \text{GCN}_{\boldsymbol{\sigma}}(\mathbf{x}_i, \mathbf{a}_i)$
            Sample $c \sim \text{Cat}(c \mid \boldsymbol{\pi})$
            Sample $\mathbf{z}_i \sim \mathcal{N}(\boldsymbol{\mu}_c|_i, \text{diag}(\boldsymbol{\sigma}_c^2|_i))$
            Obtain reconstructed $\widetilde{\mathbf{a}}_i = \tau(\mathbf{z}_i^\top \mathbf{z}_j)$ ▷ Decoder
            Compute loss, $\mathcal{L}_{\text{ELBO}}$ ▷ From (14)
                and backpropagate gradients.
        end for
    end for
    Extract community assignment $\boldsymbol{\gamma}$ via $\mathbf{z}_i$ ▷ From (18)
    Return $\boldsymbol{\gamma}, \widetilde{\mathbf{A}} = \{\widetilde{\mathbf{a}}_1, ..., \widetilde{\mathbf{a}}_N\}$

ALGORITHM 1: Variational Graph Autoencoder for Community Detection (VGAECD).

For starters, experiments are performed on datasets in accordance to Karrer and Newman. These networks (Karate and PolBlogs) are featureless and only contain structural information. The Karate network is a commonly studied empirical benchmark network for community detection. Similar to [39], only the largest connected component and its undirected form are considered for Polblogs. Next, two networks containing features are used (Cora and Pubmed) [30, 50]. Table 1 summarizes the list of datasets and their respective properties.

5.3. Baseline Methods. We establish a baseline by comparing against several state-of-the-art methods. These methods are divided into two categories. The first category comprises

discriminative methods and the second category comprises generative methods.

Discriminative Methods

(i) *Spectral Clustering* [9] is a commonly used approach for performing graph clustering. By identifying the Fiedler Vector of the Graph Laplacian, we can divide the network into two components. Repeating this process, the graph can be subdivided further, giving more clusters in the process.

(ii) *Louvain* [23] is a greedy modularity optimization method for maximizing modularity score.

TABLE 1: Empirical network datasets.

Dataset	Type	Nodes	Edges	Clusters (K)	Features
Karate	Social	34	78	2	*N/A*
Polblogs	Blogs	1,222	16,717	2	*N/A*
Cora	Citation	2,708	5,429	7	1,433
PubMed	Citation	19,717	44,338	3	500

(iii) *DeepWalk* [27], proposed by Perozzi *et al.*, is a network embedding method that performs a bias random walk on a given network.

(iv) *node2vec* [28] is a generalization of DeepWalk. It leverages on homophily and structural roles in embedding.

Generative Methods

(i) *Stochastic Blockmodel (SBM)* [38, 39] is a state-of-the-art generative model. It models the likelihood of two nodes forming an edge on the basis of stochastic equivalence. Degree Correction (D.C.) penalizes the formation of single node modules by normalizing the node degrees.

(ii) *Variational Graph Autoencoder* [15] follows the Variational Autoencoder framework by leveraging on GCN layers.

5.4. Evaluation Metrics. Some of the common approaches to evaluate detected communities are Normalized Mutual Information (NMI), Variation of Information (VI), and Modularity. In some cases, accuracy can be accurately measured (i.e., when the number of clusters K is 2). Furthermore, these measures are only possible when ground truth exists. Hence, we include other forms of measures which consider the quality of a partition without ground truths.

5.4.1. Ground Truth

(i) Accuracy measures the number of correctly classified clusters given the ground truth. Formally, given two sets of community labels, i.e., C being the ground truth and C' the detected community label, the accuracy can be calculated by

$$ACC\left(C'\right) = \frac{\sum_{i=1}^{|C|} \delta\left(c_i, c_i'\right)}{|C|} \times 100\%. \quad (19)$$

$c_i \in C$, $c_i' \in C'$, where $\delta(\cdot)$ denotes the Kronecker delta, $\delta(c_i, c_i') = 1$ when both labels match, and $|\cdot|$ denotes the cardinality of a set. For clustering tasks, accuracy is usually not emphasized as labels are known to oscillate between clusters.

(ii) NMI and VI are based on information theory. Essentially, NMI measures the "similarity" between two

community covers, while VI measures their "disimilarity" in terms of uncertainty. Correspondingly, a higher NMI indicates a better match between both covers while VI indicates the opposite. Formally [51]

$$\text{NMI}\left(C, C'\right) = \frac{2\text{I}\left(C, C'\right)}{\left(\text{H}\left(C\right) + \text{H}\left(C'\right)\right)} \quad (20)$$

and

$$\text{VI}\left(C, C'\right) = \text{H}\left(C\right) + \text{H}\left(C'\right) - 2\text{I}\left(C, C'\right), \quad (21)$$

where $\text{H}(\cdot)$ is the entropy function and $\text{I}(C, C') = \text{H}(C) + \text{H}(C') - \text{H}(C, C')$ is the mutual information function.

5.4.2. Community Quality

(i) Modularity (Q) [17] measures the quality of a particular community structure when compared to a null (random) model. Intuitively, intracommunity links are expected to be stronger than intercommunity links. Specifically,

$$Q = \frac{1}{4m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{4m} \right) \delta\left(c_i, c_j\right), \quad (22)$$

where $A_{ij} - k_i k_j / 4m$ measures the actual edge connectivity versus the expectation at random and $\delta(c_i, c_j)$ defines the Kronecker delta, where $\delta(c_i, c_j) = 1$ when both nodes i and j belong to the same community, and 0 otherwise. Essentially, Q approaches 1 when the partitions are considered good.

(ii) Conductance (CON) [52, 53] measures the separability of a community across the fraction of outgoing local volume of links in the community, which is defined as

$$\text{CON}\left(C\right) = \frac{\sum_{i \in C, j \in C'} A_{ij}}{\min\left(a\left(C\right), a\left(C'\right)\right)}, \quad (23)$$

where the nominator defines the total number of edges within community C and $a(C) = \sum_{i \in C}(j \in V)$ defines the volume of set $C \subseteq V$. A better local separability of community is achieved when the overall conductance value is the smallest.

(iii) Triangle Participation Ratio (TPR) [53] measures the fraction of triads within the community C.

$$\text{TPR}(C) = \frac{\left|\left\{v_i \in C, \ \left\{(v_j, v_k) : v_j, v_k \in C, \ (v_i, v_j), (v_j, v_k), (v_i, v_k) \in E\right\} \neq \emptyset\right\}\right|}{|C|}, \tag{24}$$

where E denotes the total number of edges in the graph G. A larger TPR value indicates a denser community structure.

5.5. Experiment Settings.

For discriminative models such as node2vec and DeepWalk, the latent representation is learned first. Next, k-means is applied to the learned latent vector with K given a priori. The parameters used for node2vec are performed using exhaustive search on variables $p, q \in \{0.25, 0.5, 1, 2, 4\}$ as suggested in [28]. Specifically, the parameters obtained were ($p = 0.5, q = 4$), ($p = 0.25, q = 0.25$), ($p = 1, q = 0.25$), and ($p = 0.5, q = 1$) for Karate, PolBlogs, Cora, and PubMed, respectively. As for DeepWalk, the parameters used are $d = 128$, $r = 10$, $l = 80$, and $k = 10$ which were the suggested values [27]. On the other hand, generative models like SBM (and D.C.) have several optimization strategies. In this case, we applied the Expectation-Maximization (EM) algorithm as suggested in [39].

For a fair comparison between VGAE and VGAECD, we used identical layer configurations for both models. The layer configurations are (32-16), (32-16), (32-8), and (32-8) for Karate, PolBlogs, Cora, and PubMed, respectively. These configurations are determined empirically as suggested in [15]. Generally, we found the first layer to be insensitive and second layer to be sensitive. By reducing the size of the second layer with respect to the number nodes we found that 8 was ideal for Cora and PubMed. The hyperparameter K is given a priori for all methods. For a fair comparison, the average of 10 runs was taken for both discriminative and generative models. All experiments were conducted on an Ubuntu 16.06 LTS machine with 64 GB of RAM and two GeForce GTX 1080 Ti graphics cards.

5.6. Experiment Results.

We first compare our result with 8 baseline methods on several state-of-the-art methods that employ unsupervised network embedding, except SBM: the only generative model that does learn a network embedding. Since VGAE is nonclustering, the two-step approach for clustering was applied, i.e., obtaining the latent vectors and subsequently applying k-means. The $*$ symbol denotes methods that were confined to structural information only.

5.6.1. Synthetic Dataset Performance.

Figure 2 depicts the performance of the proposed model in comparison to other methods. In Figure 2(a), VGAECD can be seen as a strong performer when $z_{\text{out}} \geq 4$. On the LFR benchmark graph in Figure 2(b), the performance of VGAECD is comparable to other methods. When $\mu < 0.4$, VGAECD is capable of outperforming other methods. When $\mu > 0.55$, VGAECD is seen to exhibit similar performance to other methods.

In both cases, the performance was as expected since the mixing parameter (z_{out} and μ) is consistent with the study recoverability limit in planted partitions [42, 43].

5.6.2. Empirical Dataset Performance.

Experiments performed on four different empirical datasets are shown in Tables 2, 3, 4, and 5 for Karate, PolBlogs, Cora, and PubMed, respectively. We measure the performance of clusters found using metrics as proposed in the Section 5.4 and the best values are marked in bold.

Generally, the experiments revealed that our method outperforms other methods when ground truth is given. In terms of cluster quality, VGAECD performs relatively well in terms of modularity score (Q). However, it retains competitiveness on Conductance (CON) and Triangle Participation Ratroio (TPR) measures. Since datasets such as Cora and PubMed have more than 2 clusters ($K > 2$), the accuracy of labels can be affected by label oscillation. Therefore, it is a less accurate measure for measuring cluster's label when compared to classification accuracy measures. However, accurate measures can still be obtained for datasets with only two clusters such as Karate and PolBlogs, which revealed that the proposed method is better than baseline methods. In most cases, the results of our method are comparable to SBM (D.C.). This is plausible since SBM (D.C.) has an advantage due to its prior knowledge on degree normalization. Regardless, when more than two clusters are given, the modularity score of VGAECD outperforms SBM (D.C.) as shown in Cora and PubMed datasets.

5.6.3. Time Complexity Analysis.

Since the proposed model follows the VAE framework, it employs a similar optimization method using SGVB. Therefore, it follows a linear-time complexity for one epoch, but requires L number of runs to achieve convergence. The convergence rate of NMI with respect to the number of epochs can be observed in Figure 3 in comparison to VGAE. In contrast to VGAE, the proposed method can achieve convergence at a faster rate.

5.6.4. Synthetic Network Generation.

The implication of a generative model is its ability to generate a graph when prescribed a certain set of parameters. Therefore, a synthetic network can be generated using the proposed VGAECD model. Given parameters c and ω, we can generate a network simply by following the generative process specified in Section 4.2. However, in order to vary the community structure, we can follow the planted partition's approach by including the mixing of a random network model:

$$\omega = \lambda \omega^{\text{planted}} + (1 - \lambda) \omega^{\text{random}}. \tag{25}$$

TABLE 2: Experimental results on Karate dataset.

	NMI (↑)	VI (↓)	ACC (↑)	Q (↑)	CON (↓)	TPR (↑)
Spectral Clustering	0.2297	2.0005	0.7353	0.1127	0.3702	0.7363
Louvain	0.4900	1.5205	0.3235	**0.4188**	0.2879	0.7333
DeepWalk	0.7198	0.8812	0.9353	0.3582	0.1337	0.9353
node2vec	0.8372	0.8050	0.9706	0.1639	0.4239	0.4549
Stochastic Blockmodel	0.0105	1.1032	0.4412	-0.2084	0.7154	0.4034
Stochastic Blockmodel (D.C)	0.8372	0.8050	0.9706	0.3718	**0.1282**	**0.9412**
VGAE∗ + k-means	0.6486	0.8189	0.9647	0.3669	0.1295	0.9407
VGAECD∗	**1.0000**	**0.6931**	**1.0000**	0.3582	0.1412	**0.9412**

TABLE 3: Experimental results on PolBlogs dataset.

	NMI (↑)	VI (↓)	ACC (↑)	Q (↑)	CON (↓)	TPR (↑)
Spectral Clustering	0.0014	1.1152	0.4828	-0.0578	0.5585	0.7221
Louvain	0.6446	1.0839	0.9149	0.2987	0.8130	0.1922
DeepWalk	0.7367	1.0839	0.9543	0.0980	0.3873	0.6870
node2vec	0.7545	0.8613	0.9586	0.1011	0.3827	0.6863
Stochastic Blockmodel	0.0002	1.2957	0.4905	-0.0235	0.5329	0.5657
Stochastic Blockmodel (D.C)	0.7145	0.8890	0.9496	**0.4256**	**0.0730**	**0.8101**
VGAE∗ + k-means	0.7361	0.8750	0.9552	0.4238	0.0752	0.8089
VGAECD∗	**0.7583**	**0.8583**	**0.9601**	0.4112	0.0880	0.7913

TABLE 4: Experimental results on Cora dataset.

	NMI (↑)	VI (↓)	ACC (↑)	Q (↑)	CON (↓)	TPR (↑)
Spectral Clustering	0.0651	**2.0005**	0.1252	0.0189	0.1909	**0.6196**
Louvain	0.4336	4.0978	0.0081	**0.8142**	**0.0326**	0.2821
DeepWalk	0.3796	2.7300	0.1626	0.6595	0.0396	0.4949
node2vec	0.3533	2.9947	0.1359	0.6813	0.1078	0.4902
Stochastic Blockmodel	0.0917	3.5108	**0.1639**	0.4068	0.4280	0.3376
Stochastic Blockmodel (D.C.)	0.1679	3.4547	0.1176	0.6809	0.1736	0.5112
VGAE∗ + k-means	0.2384	3.3151	0.1033	0.6911	0.1615	0.4906
VGAE + k-means	0.3173	3.1277	0.1589	0.6981	0.1517	0.5031
VGAECD∗	0.2822	3.1606	0.1532	0.6674	0.1808	0.5076
VGAECD	**0.5072**	2.7787	0.1101	0.7029	0.1371	0.4987

TABLE 5: Experimental results on PubMed dataset.

	NMI (↑)	VI (↓)	ACC (↑)	Q (↑)	CON (↓)	TPR (↑)
Spectral Clustering	0.0382	1.4341	0.3261	0.0414	0.5645	**0.4935**
Louvain	0.1983	3.6667	0.0954	**0.7726**	0.1388	0.1592
DeepWalk	0.2946	1.7865	0.3101	0.5766	**0.0499**	0.2461
node2vec	0.1197	1.9849	0.2228	0.3501	0.3170	0.2269
Stochastic Blockmodel	0.0004	1.9340	0.3080	-0.1620	0.1038	0.1965
Stochastic Blockmodel (D.C.)	0.1325	2.0035	0.3118	0.5622	0.8121	0.2441
VGAE∗ + k-means	0.2041	1.8096	0.3724	0.5273	0.1320	0.2898
VGAE + k-means	0.1981	1.8114	0.2751	0.5297	0.1283	0.2900
VGAECD∗	0.1642	1.8320	0.1956	0.4966	0.1252	0.2692
VGAECD	**0.3252**	**1.7056**	**0.4216**	0.6878	0.1636	0.4827

(a) GN benchmark graph

(b) LFR benchmark graph
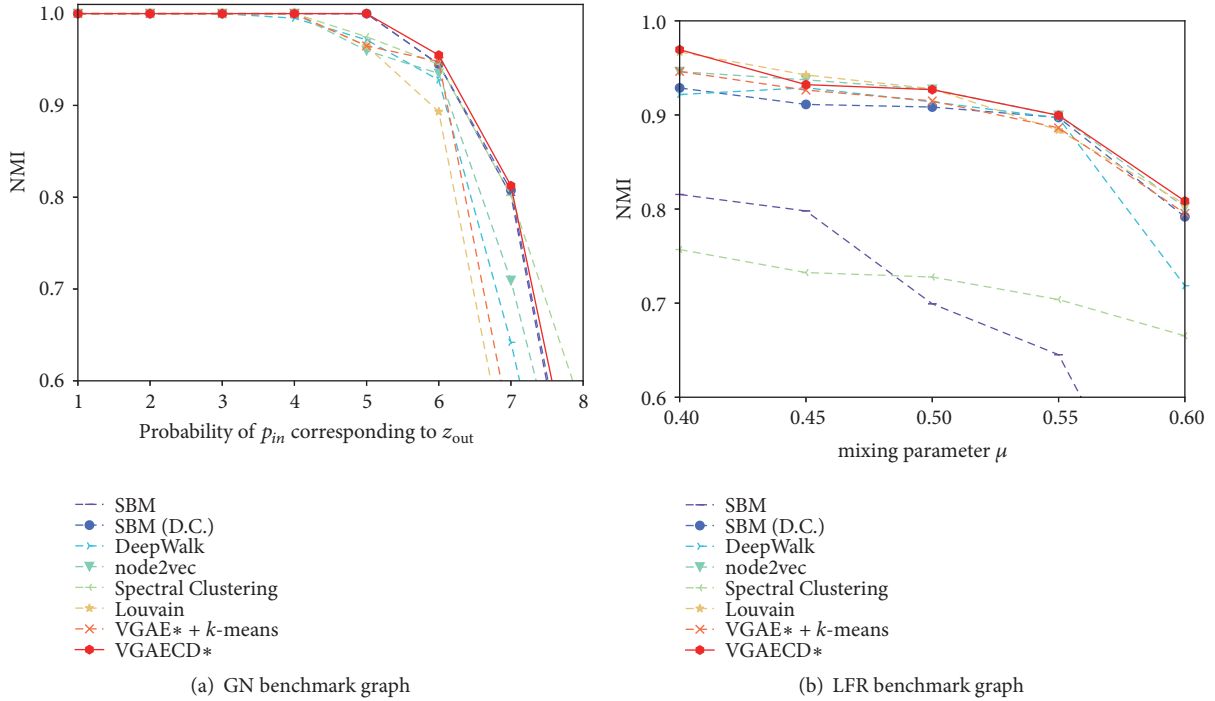
Figure 2: Comparative performance of VGAECD against other methods on synthetic networks.
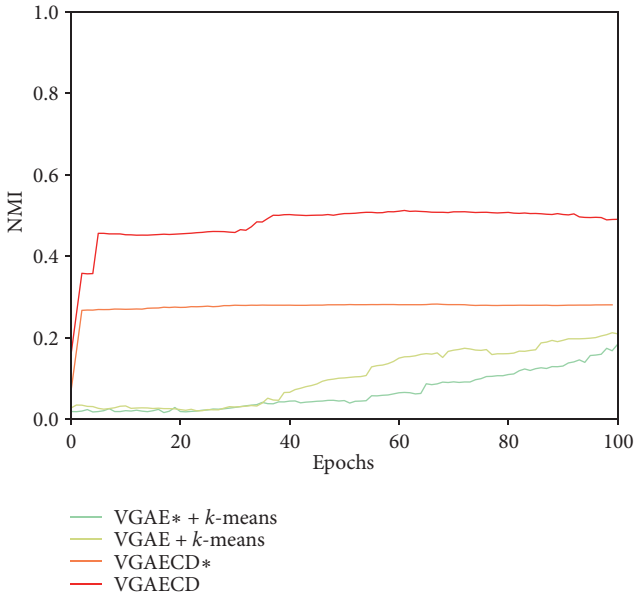
Figure 3: NMI convergence comparison.

ω^{planted} defines the amount of actual draws from the Gaussian Mixture model and ω^{random} draws from the random model. For instance, ω^{planted} can be specified as

$$\omega^{\text{planted}} = (n_1, n_2, n_3, n_4), \quad (26)$$

where n_c denotes the number of draws from the mixture model with μ_c and σ_c. In (26), we specify the number of nodes

drawn for four different communities. A generated matrix $\widetilde{\mathbf{A}}$ can be obtained as shown in decoder part of Algorithm 1. Ideally, each node is represented by $\mathbf{z}$, and the Hadamard product between $\mathbf{z}_i$ and $\mathbf{z}_j$ determines the likelihood of edge connectivity between nodes i and j which is obtained after the nonlinearity $\sigma(\cdot)$ function.

5.6.5. Network Visualization. Community assignments for Cora dataset are visualized in Figure 4. Since Cora has several disconnected nodes, only the largest connected component is visualized. Among them, VGAECD has closer resemblance to the ground truth's cluster assignment. Notably VGAECD is able to recover a community structure in the center of the network. Additionally, it also has less tendency to cluster nodes that are far away which is seen in VGAE + k-means and SBM (D.C.). DeepWalk, however, appears to have a resolution problem, resulting in larger clusters merging together. This can be seen as the number of clusters depicted in the largest component is fewer than $K = 7$. This problem is not observed in node2vec since the sampling strategies are generalization of DeepWalk. This generalization of p and q allows node2vec to explore more locations. In contrast, DeepWalk is highly restricted to visiting nodes within the pivot node's vicinity. However, to achieve the observed results, node2vec requires a costly parameter search which is not ideal. Among the baseline methods, Spectral Clustering and Louvain appear to struggle in finding a community structure, even though they performed very well on synthetic benchmark graphs. Louvain in particular had a very competitive NMI score, but visually, the results are not very satisfactory.

(a) Ground truth

(b) VGAECD

(c) Spectral clustering

(d) Louvain

(e) DeepWalk

(f) VGAE + k-means

(g) SBM (D.C.)

(h) node2vec

FIGURE 4: Visualization of community assignment on Cora Dataset (largest connected component).

6. Conclusion

In this paper, we propose a novel community detection algorithm termed Variational Graph Autoencoder for community detection (VGAECD). It generalizes VGAE for community detection tasks. The model is capable of learning both features and structural information and encodes them into a community-aware latent representation. The low-dimensional representation learned differs from previous network representation methods. Concretely, the latent representations themselves are parameters to a probabilistic graphical model, i.e., the Gaussian Mixture Model. Therefore, this allows us to draw samples from the learned model itself and generate synthetic graphs like SBM. Additionally, the flexibility of the proposed method shows that, by leveraging on more feature information, it is capable of outperforming other methods in community structure recovery. Unlike other representation learning methods which require a two-step approach (applying k-means as the second step), VGAECD is a generative model capable of recovering communities in a single step. Moreover, in comparison to existing state-of-the-art generative models such as SBM, VGAECD is community structure definition agnostic. Specifically, nodes are not forced to be similar under a specific similarity measure. This is an advantage over other community detection algorithms where the definition of community structures is always assumed. This is a desirable feature in cases where networks can have a mixture of community structures, i.e., multilayer networks.

Data Availability

All data used in our research are publicly available data. Upon request, we could point them to their respective sources.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.

[2] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8577–8582, 2006.

[3] S. Fortunato and D. Hric, "Community detection in networks: a user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.

[4] L. Tang and H. Liu, "Relational learning via latent social dimensions," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 817–826, ACM, Paris, France, July 2009.

[5] L. Tang and H. Liu, "Scalable learning of collective behavior based on sparse social dimensions," in *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp. 1107–1116, Hong Kong, China, November 2009.

[6] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the 15th IEEE International Conference on Computer Vision (ICCV '15)*, pp. 1026–1034, IEEE, Santiago, Chile, December 2015.

[8] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS '17)*, pp. 5580–5590, Long Beach, CA, USA, 2017.

[9] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On Spectral Clustering: Analysis and an Algorithm," In Advances in Neural Information Processing Systems (NIPS '02), pp. 849–856, 2002.

[10] F. Tian, B. Gao, Q. Cui, E. Chen, and T.-Y. Liu, "Learning deep representations for graph clustering," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 1293–1299, 2014.

[11] L. Yang, X. Cao, D. He, C. Wang, X. Wang, and W. Zhang, "Modularity based community detection with deep learning," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, pp. 2252–2258, 2016.

[12] P. Vincent, H. Larochelle, Y. Bengio, and P. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103, ACM, July 2008.

[13] C. Wang, S. Pan, G. Long, X. Zhu, and J. Jiang, "MGAE: marginalized graph autoencoder for graph clustering," in *Proceedings of the ACM on Conference on Information and Knowledge Management (CIKM '17)*, pp. 889–898, Singapore, November 2017.

[14] D. Chakrabarti and C. Faloutsos, "Graph mining: laws, generators, and algorithms," *ACM Computing Surveys*, vol. 38, no. 1, 2006.

[15] T. N. Kipf and M. Welling, "Variational graph auto-encoders," in *Proceedings of the Bayesian Deep Learning Workshop (NIPS '16)*, 2016.

[16] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, "Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 1965–1972, Melbourne, Australia, August 2017.

[17] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 2, Article ID 026113, 2004.

[18] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences of the United States of America* , vol. 104, no. 1, pp. 36–41, 2006.

[19] B. H. Good, Y.-A. de Montjoye, and A. Clauset, "Performance of modularity maximization in practical contexts," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 81, no. 4, Article ID 046106, pp. 1–20, 2010.

[20] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 76, no. 3, Article ID 036106, 2007.

[21] P. Pons and M. Latapy, "Computing communities in large networks using random walks," in *Proceedings of the International Symposium on Computer and Information Sciences*, pp. 284–293, Springer, 2005.

[22] M. Rosvall and C. T. Bergstrom, "Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems," *PLoS ONE*, vol. 6, no. 4, Article ID e18209, 2011.

[23] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Article ID P10008, 2008.

[24] S. Agreste, P. De Meo, G. Fiumara et al., "An empirical comparison of algorithms to find communities in directed graphs and their application in web data analytics," *IEEE Transactions on Big Data*, vol. 3, no. 3, pp. 289–306, 2017.

[25] S. Cao, W. Lu, and Q. Xu, "GraRep: learning graph representations with global structural information," in *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pp. 891–900, Melbourne, Australia, October 2015.

[26] T. Guo, S. Pan, X. Zhu, and C. Zhang, "CFOND: consensus factorization for co-clustering networked data," *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[27] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, pp. 701–710, New York, NY, USA, August 2014.

[28] A. Grover and J. Leskovec, "Node2vec: scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pp. 855–864, San Francisco, Calif, USA, August 2016.

[29] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral

filtering," in *Proceedings of the 30th Annual Conference on Neural Information Processing Systems (NIPS '16)*, pp. 3844–3852, December 2016.

[30] C. Zhuang and Q. Ma, "Dual graph convolutional networks for graph-based semi-supervised classification," in *Proceedings of the World Wide Web Conference*, pp. 499–508, Lyon, France, April 2018.

[31] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, "Kronecker graphs: an approach to modeling networks," *Journal of Machine Learning Research*, vol. 11, pp. 985–1042, 2010.

[32] A. W.-U. Ashraf, M. Budka, and K. Musial, "NetSim—The framework for complex network generator," in *Proceedings of the 22nd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES '18)*, September 2018, https://arxiv.org/abs/1805.10520.

[33] C. Seshadhri, T. G. Kolda, and A. Pinar, "Community structure and scale-free collections of Erdős-Rényi graphs," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 85, no. 5, Article ID 056109, 2012.

[34] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.

[35] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 78, no. 4, Article ID 046110, 2008.

[36] P. Brodka and T. Grecki, "mLFR Benchamark: Testing Community Detection Algorithms in Multi-layered," *Multiplex and Multiple Social Networks*, 2014.

[37] P. Bródka, *A Method for Group Extraction and Analysis in Multilayer Social Networks [PhD Thesis]*, 2012, https://arxiv.org/abs/1612.02377.

[38] T. A. Snijders and K. Nowicki, "Estimation and prediction for stochastic blockmodels for graphs with latent block structure," *Journal of Classification*, vol. 14, no. 1, pp. 75–100, 1997.

[39] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 1, Article ID 016107, 2011.

[40] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed Membership Stochastic Blockmodels," *Journal of Machine Learning Research*, vol. 9, pp. 1981–2014, 2008.

[41] D. B. Larremore, A. Clauset, and A. Z. Jacobs, "Efficiently inferring community structure in bipartite networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 90, no. 1, Article ID 012805, 2014.

[42] E. Abbe and C. Sandon, "Community detection in general stochastic block models: fundamental limits and efficient algorithms for recovery," *Foundations of Computer Science (FOCS)*, pp. 670–688, 2015.

[43] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 62, no. 1, pp. 471–487, 2016.

[44] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, and C. Zhang, "Adversarially Regularized Graph Autoencoder for Graph Embedding," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, pp. 2609–2615, Stockholm, Sweden, July 2018.

[45] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *In International Conference on Learning Representations (ICLR '14)*, 2014.

[46] A. Condon and R. M. Karp, "Algorithms for graph partitioning on the planted partition model," *Random Structures & Algorithms*, vol. 18, no. 2, pp. 116–140, 2001.

[47] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[48] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 U.S. Election: Divided they blog," in *Proceedings of the 3rd International Workshop on Link Discovery (LinkKDD '05)*, pp. 36–43, ACM, 2005.

[49] P. Sen, G. M. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad, "Collective classification in network data," *AI Magazine*, vol. 29, no. 3, pp. 93–106, 2008.

[50] Z. Yang, W. W. Cohen, and R. Salakhutdinov, "Revisiting semi-supervised learning with graph embeddings," in *Proceedings of the 33rd International Conference on Machine Learning*, pp. 40–48, 2016.

[51] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, Article ID P09008, 2005.

[52] R. Kannan, S. Vempala, and A. Vetta, "On clusterings: good, bad and spectral," *Journal of the ACM*, vol. 51, no. 3, pp. 497–515, 2004.

[53] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowledge and Information Systems*, vol. 42, no. 1, pp. 181–213, 2015.

Research Article

The Settlement Structure Is Reflected in Personal Investments: Distance-Dependent Network Modularity-Based Measurement of Regional Attractiveness

Laszlo Gadar,[1,2] Zsolt T. Kosztyan,[1,3] and Janos Abonyi [ID] **[4]**

[1]*MTA-PE Budapest Ranking Research Group, University of Pannonia, Veszprém, Hungary*
[2]*Innopod Solutions Kft., Budapest, Hungary*
[3]*Department of Quantitative Methods, University of Pannonia, Veszprém, Hungary*
[4]*MTA-PE "Lendulet" Complex Systems Monitoring Research Group, University of Pannonia, Veszprém, Hungary*

Correspondence should be addressed to Janos Abonyi; janos@abonyilab.com

How are ownership relationships distributed in the geographical space? Is physical proximity a significant factor in investment decisions? What is the impact of the capital city? How can the structure of investment patterns characterize the attractiveness and development of economic regions? To explore these issues, we analyze the network of company ownership in Hungary and determine how are connections are distributed in geographical space. Based on the calculation of the internal and external linking probabilities, we propose several measures to evaluate the attractiveness of towns and geographic regions. Community detection based on several null models indicates that modules of the network coincide with administrative regions, in which Budapest is the absolute centre, and where county centres function as hubs. Gravity model-based modularity analysis highlights that, besides the strong attraction of Budapest, geographical distance has a significant influence over the frequency of connections and the target nodes play the most significant role in link formation, which confirms that the analysis of the directed company-ownership network gives a good indication of regional attractiveness.

1. Introduction

Mining valuable information from social networks is a hard problem due to its dynamic nature [1, 2], complex structure [3, 4], and multidimensionality [5]. This paper deals with the structural issues as it tries to evaluate regional attractiveness based on a set of goal-oriented null models identified to describe the geographical distributions of company-ownership relations.

Complex multivariate socioeconomic data is widely used to monitor regional policy [6, 7]. As the usage of a different set of variables results in various rankings, the definition and selection of socioeconomic variables are the key issue in these applications. The drawback of these indicator-based approaches is that although economic behavior is socially constructed and embedded in networks of interpersonal relations [8] and strong related to location

[9], the network structure of the economy is neglected.

This paper adds a viewpoint to regional studies based on the analysis of how the network of personal investments and the founding of companies relate to the settlement hierarchy. We assume that the socially embedded economy must have a network-based imprint in the company-ownership network which is a good indication of regional attractiveness.

Attractiveness is meaningful in preferential attachment networks, where the likelihood of a new connection is proportional to degree [10] and fitness [11] of the node. These models were generalized to handle initial attractiveness [12] and latecomer nodes with a higher degree of fitness [11, 13]. It is important to note that these models generate power-law (degree) distributions that are similar to the distribution of socioeconomic variables of settlements indicating that preferential attachment is a process that can be used to

describe city grow [14–18]. In the case of geographically distributed networks, the likelihood of link formation is dependent on distance due to the cost of establishing connections and spatial constraints [19]. Connection costs also favor the formation of cliques and thus increase the clustering coefficient [20]. Space is important in social networks as most individuals connect with their spatial neighbors [20] to minimize their effort and maintain social ties [21]; e.g., the majority of our friends are in our spatial neighborhood [22]. The probability $P(d)$ that distance d separates two connected individuals is found to behave as $P(d) \sim d^{-2}$ in terms of Belgian mobile phone data [23], or generally $P(d) \sim d^{-\delta}$, as has been shown in the case of the social network of more than one million bloggers in the USA [24], in friendship network of Facebook users, and in email communication networks [25, 26].

The attractiveness of airports [27], countries for foreign investments [28], and touristic destinations [29] is evaluated based on socioeconomic variables. As many origins and destinations are present in these applications, the theory of bilateral trade flows accounts for the relative attractiveness of origin-destination pairs. The gravity model is one of the most successful empirical models in economics developed to describe such interactions across the space [30]. Almost 40 years ago, before the emergence of network science, Anderson suggested that as a force between two mass points, the number of trips from location i to location j, follows the (economic version) of the "Gravity" law, $F(d) \sim P(d) \sim I_1^{\alpha} I_2^{\alpha} d^{-\delta}$ [31]. Nowadays, many complex networks embedded in space and spatial constraints may have an effect on their connectivity patterns such as trade markets [32], migration [33], traffic flow [34], and mobile communication [23] that can be successfully modeled by a gravity model, which was also successfully applied in link prediction [35].

We assume that regions that heavily rely on local resources consist of more internal connections that form modules in networks, so the modularity of the networks which reflect socioeconomic relationships can be used to measure regional attractiveness. The goal of modularity analysis is to separate the network into groups of vertices that have fewer connections between them than inside the communities [36]. In social network analysis, community detection is a basic step in understanding the structure, function, and semantics of networks [4]. Community analysis is performed in two separate phases: first, detection of meaningful community structure from a network, and second, evaluation of the appropriateness of the detected community structure [37]. Systematic deviations from a random configuration allow us to define a quantity called modularity, that is a measure of the quality of partitions. Newman-Girvan modularity considers only the degree of nodes as a null model which is equivalent to rewiring the network whilst preserving the degree sequence [38, 39]. This random model overlooks the spatial nature of the network; thus, modules are blind to spatial anomalies and fails to uncover modules determined by factors other than mere physical proximity [19], which is the reason why several distance-dependent null models have been proposed recently [19, 37, 40, 41].

Our goal is to use the tools of network community detection to evaluate the attractiveness of the elements of settlement hierarchies (towns, statistical subregions, counties, and regions) based on their modularities as well as internal and external connection densities. We study the internal connections of the ownership network through the point of view of Newman-Girvan, spatial and gravity-based null models. As the modularity is based on the difference between the actual and evaluated values of weight of edges, the real spatial network more accurately describes the null model, and the total modularity tends to be zero, so the modules highlight the hidden structural similarities. We developed a visualization technique to analyze these unknown effects on community structure which can explain the attractiveness of a settlement/region. Besides measuring the attractiveness, we utilize the Louvain community detection algorithm [42, 43] to identify closely related regions. We examine the complete investment network of Hungarian companies to explore how the ownership connections are geographically distributed, what is the structure of the network, and what are the common connection directions, as well as how the extracted information is correlated to the settlement hierarchy. The studied database contains information about the owners and addresses of the companies. The results highlight the fact that distance dependence of the investment connections is more significant than was found in online social networks [22, 26, 44]. The analysis shows that the network is hierarchical and modular as well as shaped according to the settlement hierarchy, in which Budapest is the absolute center, and the centers of counties function as hubs.

The outline of this paper is as follows: Section 2.1 presents the company-ownership network. The metrics related to attractiveness are given in the Appendix. Section 2.2 describes the null models designed by us to measure modularity as well as handling physical proximity and presents how closely related regions can be explored based on the modularity-related merging of towns and subregions. The results and discussion are provided in Section 3.

2. Problem Formulation: Settlement Hierarchy and Community Structure in Personal Investment Patterns

2.1. Network Representation of Personal Investment Patterns.
The proposed methodology is based on the analysis of a directed investment network represented by an asymmetric biadjacency matrix $\mathbf{A}^{[p,co]}$, whose elements are defined as

$$a_{i,j}^{[p,co]} = \begin{cases} 1 & \text{if the } i\text{-th person owns the } j\text{-th company} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

As the addresses of the owners and their companies are known, connections between companies and their owners define ties between geographic locations.

According to the levels of the settlement hierarchy, a four-level study can be defined to describe how towns, regions,
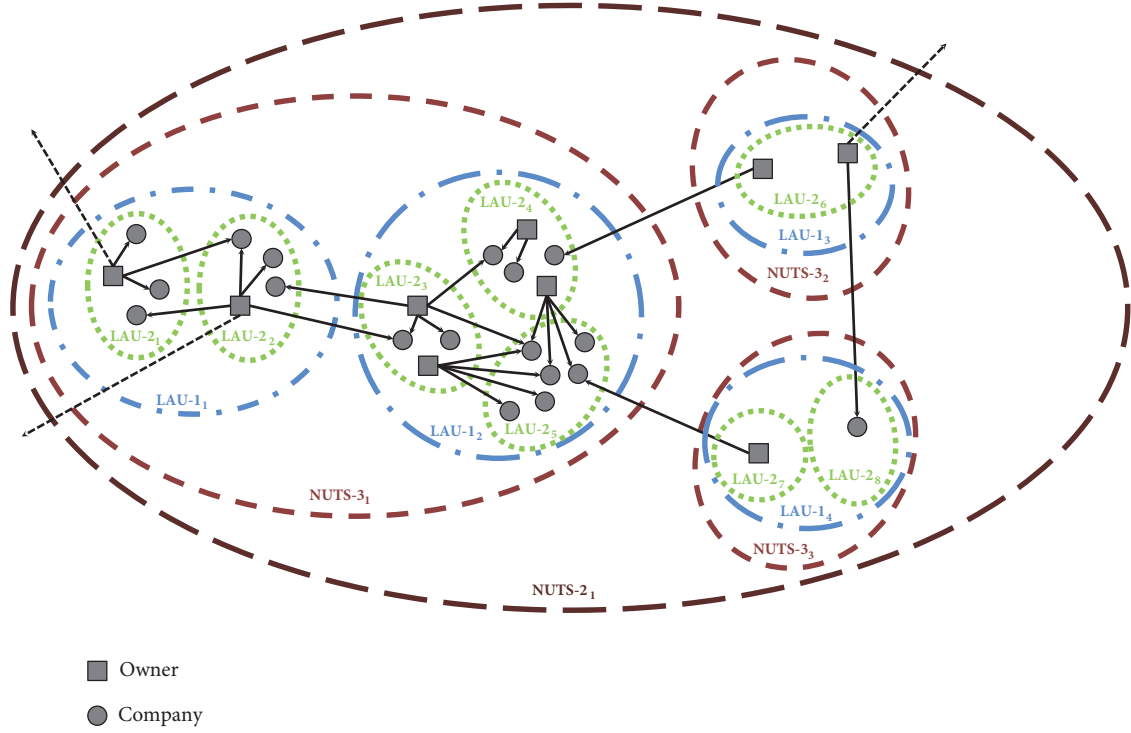
Figure 1: Company-ownership relations connect the elements of the settlement hierarchy (Settlement (LAU 2), statistical subregion (LAU 1), small-region (NUTS 3), and region (NUTS 2)).

or counties are connected through company ownerships (see Figure 1). Although companies also own shares in other companies, as we intended to study the attractiveness of economic regions based on personal investment decisions, we examined only companies that belong to individuals.

The levels of the settlement hierarchy [l] are defined based on the nomenclature of territorial units for statistics classification (NUTS) and the two levels of local administrative units (LAUs):

$$
l = \begin{cases}
1 & \text{town/settlement - LAU 2, formally NUTS 5 level} \\
2 & \text{statistical sub-region - LAU 1, formally NUTS 4 level} \\
3 & \text{small regions/counties, NUTS 3 level} \\
4 & \text{regions of regional policies, NUTS 2 level}
\end{cases} \quad (2)
$$

(Please note that, for simplicity, the term "town" is used for all cities and villages.)

People and their companies are assigned to geographic regions by the $\mathbf{A}^{[co,l]}$ and $\mathbf{A}^{[p,l]}$ incidence matrices, whose elements are defined as follows:

(i) $a_{i,j}^{[co,l]}$ with element one if the headquarter of the i-th company is situated in the j-th geographic region at the level l of the settlement hierarchy,

(ii) $a_{i,j}^{[p,l]}$ with element one if the i-th person is situated in the j-th geographic region at the level l of the settlement hierarchy,

so the directed weighted network that defines the number of investment connections between the regions can be defined as

$$
\mathbf{A}^{[l]} = \left(\mathbf{A}^{[p,l]} \right)^T \times \mathbf{A}^{[p,co]} \times \mathbf{A}^{[co,l]}. \quad (3)
$$

Although companies may have many local divisions, the links between the towns are defined only by connecting the permanent addresses of the owners and the location of the headquarter. This arrangement results in a transparent and easily interpretable network as people and companies are assigned to only one location. The resultant network describes how investments unite the locations; e.g., the adjacency matrix $\mathbf{A}^{[1]}$ defines the number of links between the towns, and the degrees of the nodes represent the number of incoming and outgoing investments to the j-th and from the i-th town, respectively:

$$
k_j^{[l,in]} = \sum_i a_{i,j}^{[l]} \quad (4)
$$

$$
k_i^{[l,out]} = \sum_j a_{i,j}^{[l]}. \quad (5)
$$

The total number of ownership relationships is equal to the sum of the edge weights of the networks:

$$
L = \sum_i \sum_j a_{i,j}^{[l]}, \quad \forall l, \quad (6)
$$

where i and j represent the indices of the geographic regions at the level l of the settlement hierarchy.

It should be noted that as L represents the total number of connections, its value is independent of at which hierarchy level the edge weights are summarised.

Similarly, the total number of companies and investors can be calculated by summing the number of companies and people at any hierarchy level, respectively:

$$N^{[co]} = \sum_{j=1} n_j^{[l,co]},$$

$$N^{[p]} = \sum_{j=1} n_j^{[l,p]}, \qquad (7)$$

$$\forall l,$$

where j represents the index of the geographic regions at the level l of the settlement hierarchy.

As people and companies are assigned only to one geographical region with the $\mathbf{A}^{[co,l]}$ and $\mathbf{A}^{[p,l]}$ incidence matrices, the number of people and companies at the j-th region of the $[l]$-th level of the settlement hierarchy can be calculated as

$$n_j^{[l,co]} = \sum_i a_{i,j}^{[co,l]} \qquad (8)$$

$$n_j^{[l,p]} = \sum_i a_{i,j}^{[p,l]}. \qquad (9)$$

The number of internal and external links of the network and the analysis of the local densities can be used to measure the attractiveness of the regions (see the Appendix). The following main body of the paper focuses on models that can be used to explore the communities in the network.

2.2. Evaluation of the Community Structure in the Settlement Hierarchy.

The key idea of the methodology is that geographical regions can be interpreted as nonoverlapping communities of investors and companies as they belong to exactly one region among the set of these regions on the l-th level of the hierarchy, $C^{[l]} = \{C_1^{[l]}, C_2^{[l]}, \ldots, C_l^{[l]}, \ldots, C_{n_{c,nk}}^{[l]}\}$.

From the view of a community, the external degree is the number of links that connect the i-th community to the rest of the network, while the internal degree is the number of links between companies and owners in the same community, in other words, at the same location at the l-th level of the hierarchy (for more details see Appendix A). Recently, a wide variety of $f(C)$ metrics have been proposed to evaluate the quality of communities on the basis of the connectivity of their nodes [37]. The following subsections will demonstrate how these metrics can be interpreted to evaluate the attractiveness of geographical regions.

2.2.1. Modularity of a Region and Level of a Settlement Hierarchy.

Classical modularity optimization-based community detection methods utilize $f(C)$ metrics that are based on the difference between the internal number of edges and their expected number [39, 45]:

$$f(C) = (\text{fraction of edges within communities})$$
$$- (\text{expected fraction of such edges}). \qquad (10)$$

In the case of the proposed directed network, this difference can be formulated as

$$f\left(C^{[l]}\right) = \frac{1}{L} \sum_{i,j} \left(a_{i,j}^{[1]} - p_{i,j}^{[1]}\right) \delta\left(C_i^{[l]}, C_j^{[l]}\right), \qquad (11)$$

where $p_{i,j}^{[1]}$ represents the number of estimated investments proceeding from the i-th to the j-th town and $\delta(C_i^{[l]}, C_j^{[l]})$ is the Kronecker delta function that is equal to one, if the i-th and j-th towns are assigned to the same region on the l-th level of the hierarchy (e.g., $\delta(C_A^{[2]}, C_B^{[2]}) = 1$ when towns A and B are situated in the same statistical subregion).

The modularity of the partition $C^{[l]}$ can be calculated as the sum of the modularities of the $C_c^{[l]}, c = 1, \ldots, n_c^{[l]}$ communities:

$$M_c^{[l]} = \frac{1}{L} \sum_{(i,j) \in C_c^{[l]}} \left(a_{i,j}^{[1]} - p_{i,j}^{[1]}\right). \qquad (12)$$

The value of the modularity $M_c^{[l]}$ of a cluster/region $C_c^{[l]}$ can be positive, negative, or zero. Should it be equal to zero, the community has as many links as the null model predicts. When the modularity is positive, then the $C_c^{[l]}$ subgraph tends to be a community that exhibits a stronger degree of internal cohesion than the model predicts.

Using the proposed matrix representation, the calculation of the internal links at a given level of the hierarchy is straightforward, so the modularity can be easily calculated based on the diagonal elements of the adjacency matrices of the network and its null model:

$$f\left(C^{[l]}\right) = \sum_{c=1}^{n_c^{[l]}} M_c^{[l]} = \frac{1}{L} \sum_c a_{c,c}^{[l]} - \frac{1}{L} \sum_c p_{c,c}^{[l]}, \qquad (13)$$

where $a_{c,c}^{[l]}$ represents the number of internal links in the c-th community/region on the l-th hierarchy level while $p_{c,c}^{[l]}$ is the expected number of these internal links calculated by the null model.

2.2.2. Null Models for Representing Regional Attractiveness.

The critical element of the methodology is how the $p_{i,j}^{[1]}$ connection probabilities of the towns are calculated. The most widely applied *null model* is the random configuration model which calculates the edge probabilities assuming a random graph conditioned to preserve the degree sequence of the original network:

$$p_{i,j}^{[1]} = \frac{k_i^{[1,out]} k_j^{[1,in]}}{L}. \qquad (14)$$

This randomized null model is inaccurate in most real-world networks [41].

As we measure the attractiveness of the regions based on the probability of link formation, it is beneficial to utilize attractiveness-related variables in the model as well as taking the distance-dependent link structure into account. Firstly,

we generalize the model by defining the node importance measures I_i^{out} and I_j^{in}:

$$p_{i,j}^{[1]} = \gamma I_i^{out} I_j^{in}. \tag{15}$$

As is expected from the null model, to fulfill the following equality,

$$\sum_{i,j} p_{i,j}^{[1]} = \sum_{i,j} a_{i,j}^{[1]} = L, \tag{16}$$

the importance measures are normalized as $\sum_i I_i^{out} = 1$ and $\sum_j I_j^{in} = 1$:

$$I_i^{out} = \frac{x_i^{\alpha}}{\sum_j x_j^{\alpha}},$$
$$I_j^{in} = \frac{x_j^{\beta}}{\sum_i x_i^{\beta}}, \tag{17}$$

where the parameters $\alpha, \beta > 0$ reflect the importance of the x_i and x_j variables used to express the probability of forming an edge from the i-th to the j-th node. Please note that when $\alpha = 1$ and $\beta = 1$, $x_i = k_i^{[1,out]}$, $x_j = k_j^{[1,in]}$, and $\gamma = L$, the model is identical to the random configuration model of a weighted directed graph.

To model the probability of distance-dependent link formation, the model defined by (15) is extended by a deterrence function $f(d_{i,j})$ which describes the effect of space [20]:

$$p_{i,j}^{[1]} = \gamma I_i^{out} I_j^{in} f\left(d_{i,j}\right). \tag{18}$$

The function $f(d_{i,j})$ can be directly measured from the data by a binning procedure similar to that used in [19]:

$$f(d) = \frac{\sum_{i,j|d_{i,j}=d} a_{i,j}^{[1]}}{\sum_{i,j|d_{i,j}=d} I_i^{out} I_j^{in}} \tag{19}$$

whose function is proportional to the weighted average of probability $(1/\gamma)\, a_{i,j}^{[1]}/(I_i^{out} I_j^{in})$ of a link existing at distance d.

When the distance dependence of the connection probability is handled by an explicit function, various modifications of the gravity law-based configuration model can be defined: $f(d) = 1/d_{i,j}^{\delta}$ [34, 46], $f(d) = \exp(-d_{i,j}/\delta)$ [47], or $f(d) = d_{i,j}^{-\delta} \exp(-d_{i,j}/\kappa)$ [48].

To ensure that the sum of the expected number of links is equal to L (see (16)), in this distance-dependent model γ should be normalized as

$$\gamma = \frac{L}{\sum_{i,j} I_i^{out} I_j^{in} f\left(d_{i,j}\right)}. \tag{20}$$

Several models can be defined based on what kind of indicators are selected in the model. When the nodes are considered to be equally important, in other words, $I_i = I_j = 1$, only the distance determine the link formation probability,

$f(d_{i,j})$. The importance of the nodes can be interpreted as the number of investors and companies, so $I_i = (n_i^{[l,p]})^{\alpha}$ and $I_j = (n_j^{[l,co]})^{\beta}$. The null model can be defined based on the random configuration model, which results in the selection of the variables as $I_i = (k_i^{[l,out]})^{\alpha}$ and $I_j = (k_j^{[l,in]})^{\beta}$. Finally, socioeconomic indicators, like the number of inhabitants, or their complex combinations can be utilized.

When $f(d) = 1/d_{i,j}^{\delta}$, the parameters α, β, δ can be estimated as a regression problem. The identified parameters indicate the sensitivity, i.e., importance, of the variables that can be sorted by their importance as suggested in classical gravity law-based studies, like in [20].

2.2.3. Economic Relations of the Regions. Connections that interlink communities indicate their relationships and possibilities to merge modules/regions that are strongly connected. We combine regions and determine the gain of the merged modularity in a similar way to the Louvain community detection algorithm [42]. The $\Delta M_{i,j}$ modularity change obtained by merging the i-th and j-th communities can be calculated as the difference between the actual and predicted number of interlinking nodes:

$$\Delta M_{i,j}^{[l]} = \frac{1}{L}\left(a_{i,j}^{[l]} - p_{i,j}^{[l]}\right) + \frac{1}{L}\left(a_{j,i}^{[l]} - p_{j,i}^{[l]}\right). \tag{21}$$

The resultant symmetric modularity gain matrix can be calculated as

$$\Delta \mathbf{M}^{[l]} = \left(\mathbf{B}^{[l]}\right)^T + \mathbf{B}^{[l]}, \tag{22}$$

where $\mathbf{B}^{[l]} = \mathbf{A}^{[l]} - \mathbf{P}^{[l]}$ is the so-called modularity matrix [38].

The Louvain algorithm moves a node i in the community for which the gain in modularity is the largest. If no positive gain occurs, i remains in its original community. After merging the nodes/regions, a new network is constructed whose nodes are in the communities identified earlier. This method can be used to explore regions (modules) formed by the elements of the l-th settlement hierarchy with different null models. Although model-based communities can be identified by this approach and compared to regions of a larger hierarchy level as modules of ground truth, the main goal of the analysis of $\mathbf{M}^{[l]}$ is to measure the strength of relationships between the regions.

The following section demonstrates the applicability of the previously presented toolset in the analysis of the network of Hungarian companies.

3. Results and Discussion

3.1. Description of the Studied Dataset. The studied dataset represents $L = 1,077,090$ ownership relations between $N^{[p]} = 531,249$ people and $N^{[co]} = 868,591$ Hungarian companies in 2013. It should be noted that only less than 10% of the ownership connections are defined based on how companies possess shares in other companies, so, although only personal investments are studied, the results reflect the attractiveness

TABLE 1: Number of edges inside the settlement hierarchies.

	Town-level	sub-Region-level	County-level	Region-level
Number of nodes, N	3,111	175	20	7
Number of internal ties	797,492	846,309	893,559	969,995
Number of external ties	279,598	230,781	183,531	107,095
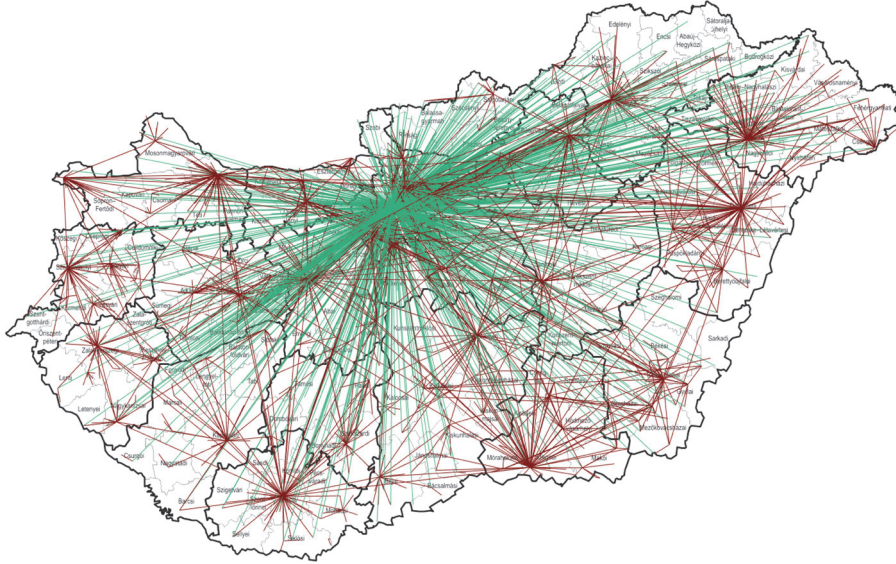
FIGURE 2: Map of the town-level company-ownership network. Edges with more than 10 ownership connections are shown. Edges connected to the capital (Budapest) are denoted by green lines.

of the towns and regions as the generated network covers more than 90% of the investment-type connections.

The owners and companies were assigned to settlements, and the related settlement hierarchy covers 3,155 towns (level LAU 2, formally level NUTS 5), 175 statistical subregions (level LAU 1, formally level NUTS 4), 20 small regions/counties in level NUTS 3, and 7 regions in level NUTS 2.

74% of the connections remain within the borders of the towns, which also reflects the high degree of modularity of the network (for more details, see Table 1). 302, 781 connections are within Budapest and 45, 559 connections point out of the city, while 89, 944 connections point into the capital. The map of the regional connections between the people and companies can be generated using the obtained connectivity matrix and the latitudes and longitudes of the towns (see Figure 2). It can be seen that the network reveals a hierarchical and modular structure reflecting that the Hungarian economy is concentrated around the capitals of the counties and Budapest, the capital of the country. The majority of the companies are situated in these locations; consequently, the network follows the structure of online social networks [44]; in other words, it is also structured according to the settlement hierarchy, in which Budapest is the absolute center of the network and the centers of counties also function as hubs.

3.2. Measuring Attractiveness. The densities inside towns and regions can highlight the modular structure of the
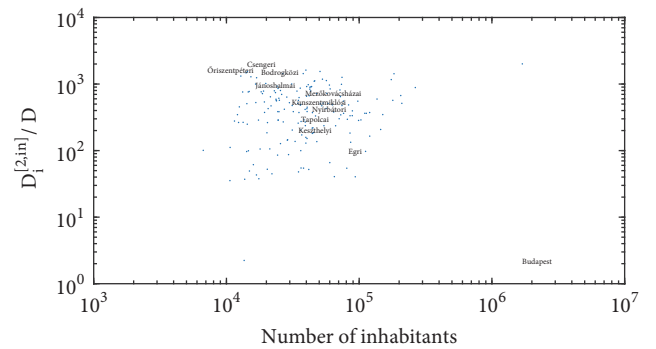
FIGURE 3: Network density as a function of the number of inhabitants on the level LAU 1.

company-ownership network. As shown in Figure 3, these densities are significantly higher in most subregions and a negative correlation exists between the size of the regions and the number of their inner connections ($r = 0.298$, $p < 10^{-4}$). As illustrated by the results, smaller locations are much more isolated than larger ones, like Budapest. The same result is obtained by the analysis of the external density-based openness measure which we consider as a main measure of attractiveness (see Appendix A for more details). As shown in Figure 4, bigger regions exhibit lager openness values reflecting their higher degree of attractiveness ($r = 0.94$, $p < 10^{-10}$).

TABLE 2: Performances of distance-dependent null models.

Nodes/Null models	p^{spa}	$p^{\alpha,\beta}$	p^{grav}
$I_i^{out} = I_j^{in} = 1$	0.28100	0.28113	0.28093
$I_i^{out} = N^{[p]}$, $I_j^{in} = N^{[co]}$	0.08915	0.01359	0.00651
$I_i^{out} = k_i^{[1,out]}$, $I_j^{in} = k_j^{[1,in]}$	0.05759	0.01389	0.00642
$I_i^{out} = \text{Inhabitants}_i$, $I_j^{in} = \text{Inhabitants}_j$	0.12106	0.01456	0.00650
$I_i^{out} = \text{TDI}_i$, $I_j^{in} = \text{TDI}_j$	0.07142	0.01482	0.00644
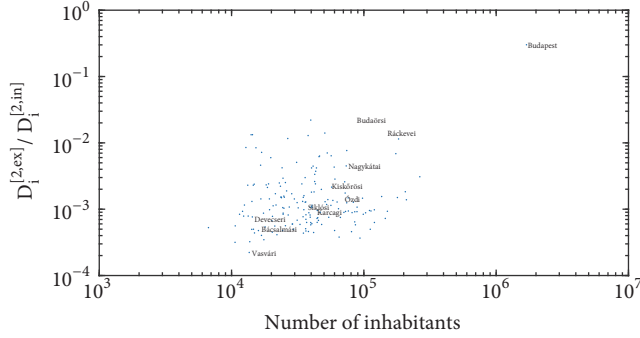
FIGURE 4: Openness of small regions (LAU 1 level) as a function of the number of their inhabitants.
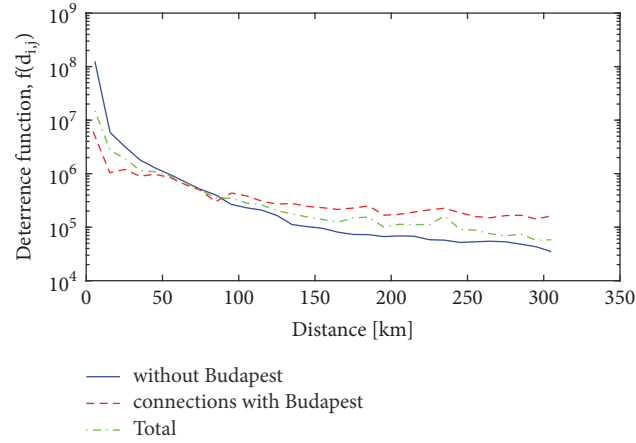
FIGURE 5: Empirically derived deterrence function determined by (19), where $I_i^{[in]} = n_i^{[1,p]}$, $I_j^{[in]} = n_j^{[1,co]}$.

3.3. The Effect of Geographical Distance. To address the effect of distance decay on link formation, the observed ties between the towns were compared with their expected number calculated from a probabilistic model.

A resolution of 10 km was used for binning the distance distribution (see Figure 5). The exponent of distance decay according to our data is -1.1057. It should be noted that the effect of the capital city is so high, the probability of forming connections with Budapest is slightly less distance-dependent, and the exponent of distance decay with regard to these connections is only -0.6385.

The distance-dependent link formation probability can be explained by the notion that the costs of establishing and maintaining the connections are also distance-dependent. This assumption can be confirmed by the fact that the distance has a much stronger effect on investment ties than on online social networks in Hungary (where the exponent of distance decay is -0.6) [44], probably since the cost of keeping connections is less dependent on distance than the management of a company far from the permanent address of the owner.

3.4. Comparison of the Null Models. Based on the utilized distance function, three different types of models can be defined. When $f(d)$ is a deterrence function defined by (19), the models are denoted as $p^{spa} = \gamma I_i^{out} I_j^{in} f(d)$. $p^{\alpha,\beta} = \gamma (I_i^{out})^\alpha (I_j^{in})^\beta f(d)$ represents the parametric version of this model, when the exponents α and β are optimized to achieve a more accurate approximation of connections between towns. $p_{i,j}^{grav} = \gamma (I_i^{out})^\alpha (I_j^{in})^\beta / d^\delta$ represents the gravity-type models.

Five sets of I_i^{out}, I_j^{in} variables were defined, including simple metrics like the numbers of nodes and edges [1] in addition to socioeconomic variables, like the number of inhabitants and Total Domestic Income (total income received by all sectors of the economy including the sum of all wages, profits, and taxes, minus subsidies). Based on the combination of different variables and distance functions, 15 different models were identified:

$$\min_{\alpha,\beta,\gamma} E_m (\alpha, \beta, \gamma) = \frac{1}{L} \left\| \mathbf{A}^{[1]} - \mathbf{P}^{[1]} \right\|_2 . \tag{23}$$

As summarized in Table 2, by taking the distance into account, the accuracy of the model is significantly improved. Among distance-dependent models, the gravity models perform best (in comparison, the accuracy of the distance independent random configuration model is 0.16494).

The Total Domestic Income (TDI) is one of the best indicators. The identified α, β, and δ parameters reflect the importance of the I_i^{out}, I_j^{in}, and d variables in the models (e.g., in the case where $I_j^{in} = TDI_j$ and $I_i^{out} = TDI_i$, the resultant nonlinear regression model is $p_{i,j} = 0.12 \cdot ((I_i^{out})^{0.37} \cdot (I_j^{in})^{0.81})/d^{1.58}$ (see Table 3)), which can be interpreted as the notion that the number of connections between location i and location j is increased by 0.37% as a result of 1.0% growth of TDI in location i. Similarly, the number of connections between location i and location j is increased by 0.81% as a result of 1.0% growth of TDI in location j. According to the gravity-type models, the importance of the target/destination

TABLE 3: Coefficients of the parametric models that reflect the importance of the variables.

Nodes/Parameters	$p^{\alpha,\beta} = \gamma \left(I_i^{out}\right)^{\alpha} \left(I_j^{in}\right)^{\beta} f(d)$		$p_{i,j}^{grav} = \gamma \left(I_i^{out}\right)^{\alpha} \left(I_j^{in}\right)^{\beta} /d^{\delta}$		
	α	β	α	β	δ
$I_i^{out} = N^{[p]}, I_j^{in} = N^{[co]}$	1.08373	0.91787	0.34984	0.67191	1.63711
$I_i^{out} = k_i^{[1,out]}, I_j^{in} = k_j^{[1,in]}$	1.05439	0.94455	0.35652	0.69045	1.59439
$I_i^{out} = \text{Inhabitants}_i, I_j^{in} = \text{Inhabitants}_j$	0.99347	1.15642	0.40654	0.88313	1.52391
$I_i^{out} = TDI_i, I_j^{in} = TDI_j$	0.98571	1.03669	0.37367	0.81425	1.58060

FIGURE 6: Comparison between the number of the edge weights $a_{i,j}^{[1]}$ and their estimated values $p_{i,j}^{[1]}$ generated by different null models on the town level (LAU 2) settlement hierarchy when $I_i^{out} = k_i^{[1,out]}$ and $I_j^{in} = k_j^{[1,in]}$. The + symbols represent the inner connections that form a separate cluster. This plot directly reflects the goodness of fit as the model estimates the connections of the towns.

locations (β) is greater than the importance of the sources (α) regardless of how the strengths of the nodes are interpreted.

3.5. Evaluation of the Modularities.

As modularity-based community detection evaluates the set of $a_{i,j}^{[1]} > p_{i,j}^{[1]}$ edges (and the related nodes) whose weights are underestimated by the null model (see (11)), we designed a plot that compares $a_{i,j}^{[1]}$ with $p_{i,j}^{[1]}$ to highlight the set of potential edges that can be used to form communities.

Four null models based on the $I_1 := k_i^{[1,out]}$ and $I_2 := k_j^{[1,in]}$ Newman and Girvan model are compared in Figure 6. In all models, the inner connections (represented by +) form a separate cluster which confirms that 74% of the connections remain within the borders of the towns. The first model (p^{NG}) shows that more inner connections exist than would be expected based on the random configuration network. The spatial models p^{Spat} and $p^{\alpha,\beta}$ handle the dependence on distance of the connections, so a slightly smaller difference is shown in the number of the experienced and expected inner connections. It is reflected in Figure 7 that during the aggregation procedure the qualitative behavior of the models does not change.

The difference between the expected number of interconnections is higher in the case of smaller settlements which indicates that small regions are not as attractive as would be expected from their number of nodes. The gravity model p^{Grav} well estimates the inner connections thanks to the exponents $\alpha = 0.35652$ and $\beta = 0.69045$ whose parameters effectively represent that the increase in the number of connections affects the attractiveness in a nonlinear fashion. This phenomenon is much more interesting when the utilized variables can be interpreted as economic potentials. When TDI is applied in the gravity model, $\alpha = 0.37367$ and $\beta = 0.81425$. These values and Figure 8 confirm that gravity-based models behave similarly and, therefore, reflect the same mechanism of attractiveness.

3.6. Forming Communities.

Connections that interlink communities are indicative of their relationships. The effect of these interlinks can be studied by the change in modularity (see (21)) expressed as $\Delta \mathbf{M}^{[l]} = (\mathbf{B}^{[l]})^T + \mathbf{B}^{[l]}$.

To determine the community structure, the MATLAB implementation [49] of the greedy Louvain algorithm [50] was used. Towns and subregions were used as an initial
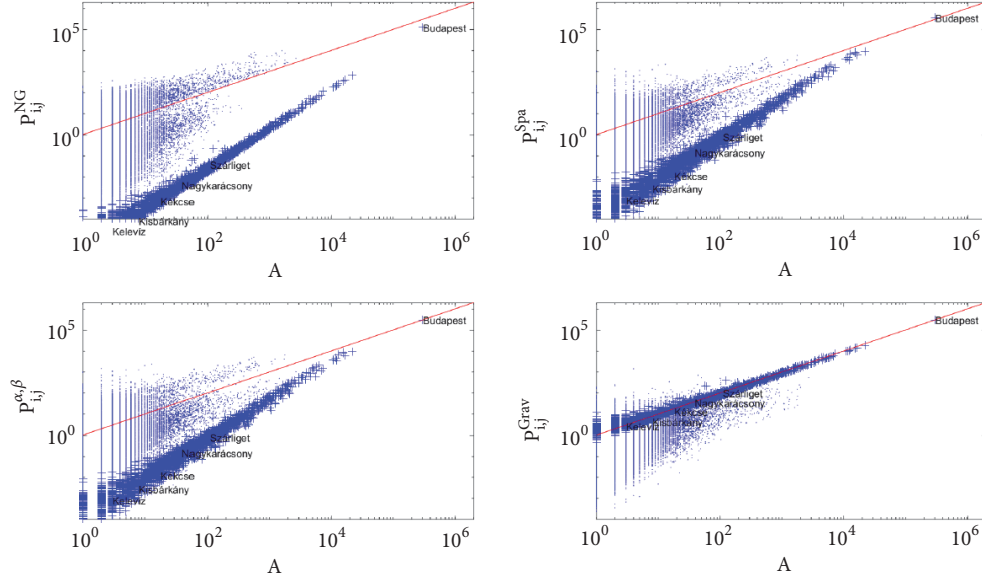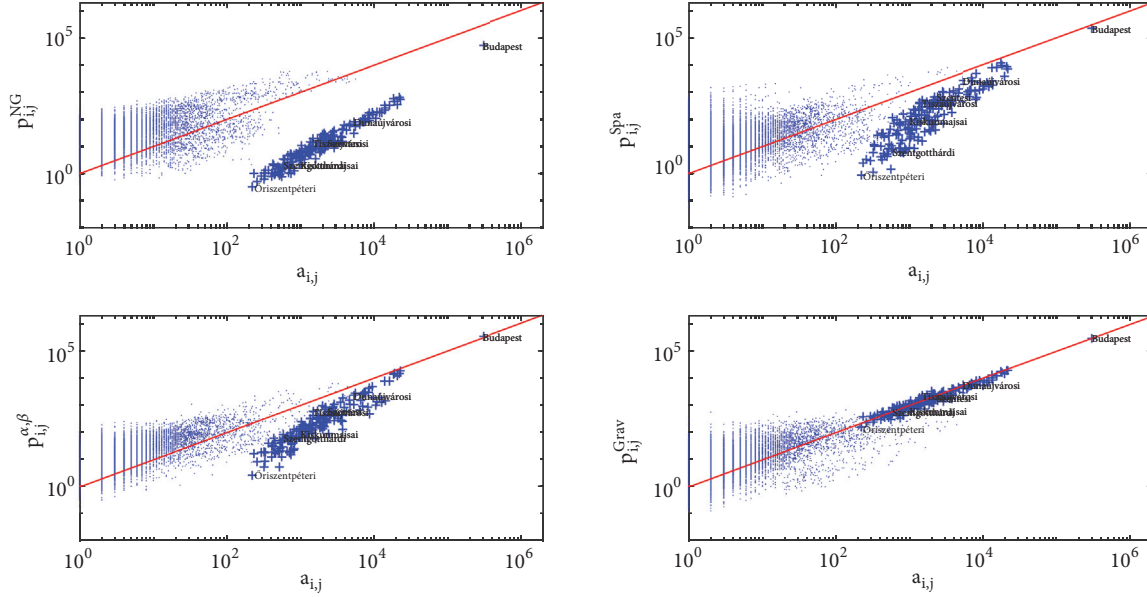
FIGURE 7: Comparison between the number of the edge weights $a_{i,j}^{[2]}$ and their estimated values $p_{i,j}^{[2]}$ generated by different null models at level LAU 1 of the settlement hierarchy when $I_i^{out} = k_i^{[1,out]}$ and $I_j^{in} = k_j^{[1,in]}$. The + symbols represent the inner connections that form a separate cluster. This plot reflects that, during the aggregation procedure, the qualitative behavior of the models does not change; furthermore, the same phenomena can be observed as in Figure 6.
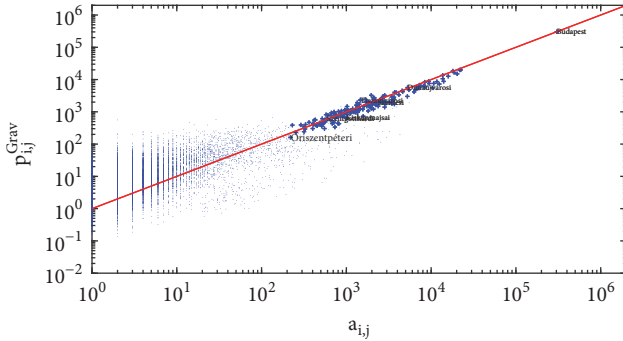
FIGURE 8: Comparison between the number of the edge weights $a_{i,j}^{[2]}$ and their estimated values $p_{i,j}^{[2]}$ generated by the gravity null model at level LAU 1 of the settlement hierarchy when $I_i^{out} = TDI_i$ and $I_j^{in} = TDI_j$. The + symbols represent the inner connections that form a separate cluster.

partition $\mathbf{B}^{[l]}$. As shown in Figure 9, the community structure formed based on the null model p^{NG} almost perfectly reconstructs the counties confirming that the settlement structure is reflected in terms of the personal investments.

Different null models provide different viewpoints with regard to community detection. The NG null model does not handle the distance dependence of the connections so the matrix $\mathbf{B}^{[l]} = \mathbf{A}^{[l]} - \mathbf{P}^{[l]}$ of the modeling errors reflects the distance dependence of the connections. Therefore, the resulting communities form spatial clusters. On the contrary, communities formed by the gravitational models

reflect distance-dependent differences less. According to the resultant maps, the attractiveness of Budapest is highlighted as only small since closed regions were not assigned to the module of the capital (see Figure 10(a)). It is interesting to note that all the centers of counties were assigned to the community of Budapest in gravitational model which also confirms the hierarchical structure of the network. To highlight the hierarchical structure and increase the sensitivity of the model, a resolution parameter was introduced into the model (see Appendix B) that can be adapted to detect similar region-pairs as shown in Figure 10(b).

Communities formed with the NG null model (see Figure 9) and the TDI-based gravity models (see Figure 10) significantly differ. The interpretation of the communities and these differences should rely on the understanding of the concept of the modularity. The utilised modularity detection algorithm generates partitions in which the links are more abundant within communities than would be expected from the employed model.

As the NG null model only uses the basic structural information encoded in the adjacency matrix, when the probabilities of the connections are dependent on distance, the resulting communities will represent closer geographical regions. As Table 1 and Figures 6 and 7 show, most of the connections remain within the county borders, so it is natural that the resultant 30 communities are almost identical to the counties.

Since the Hungarian road network reflects the administrative regions, it can be shown that the distance strongly affects the probability of the connections. This distance dependence of the connection probability can be incorporated into the null model by the proposed gravity model.
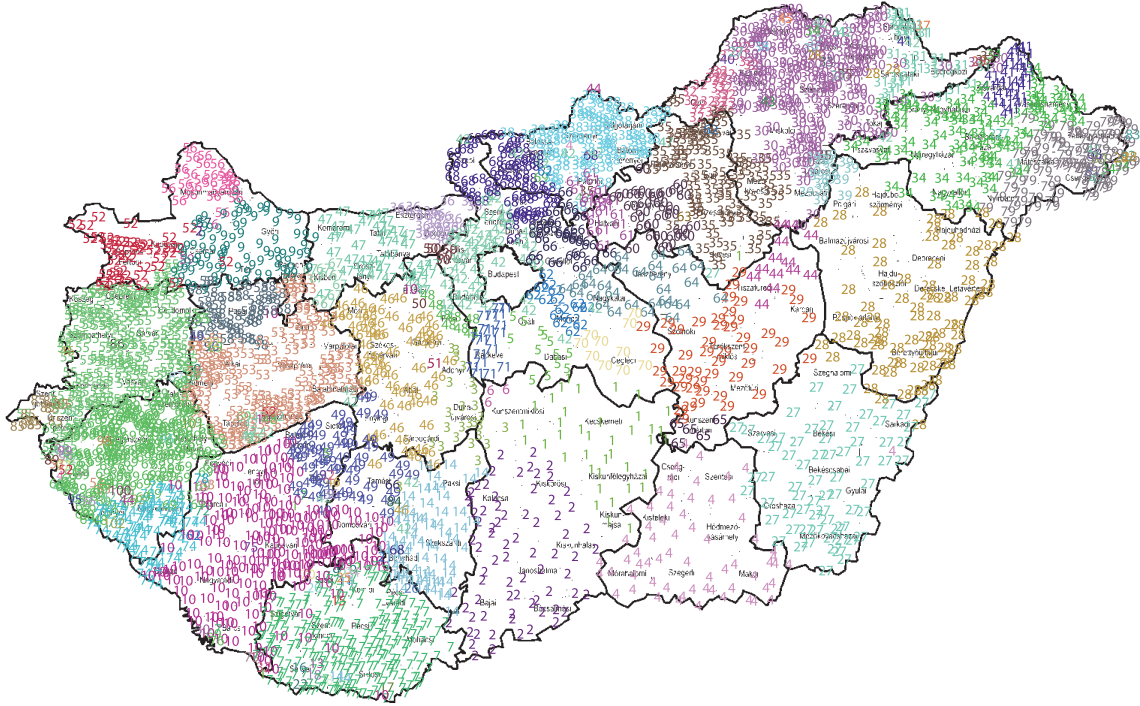
(a) Initial nodes are towns ($l = 1$)
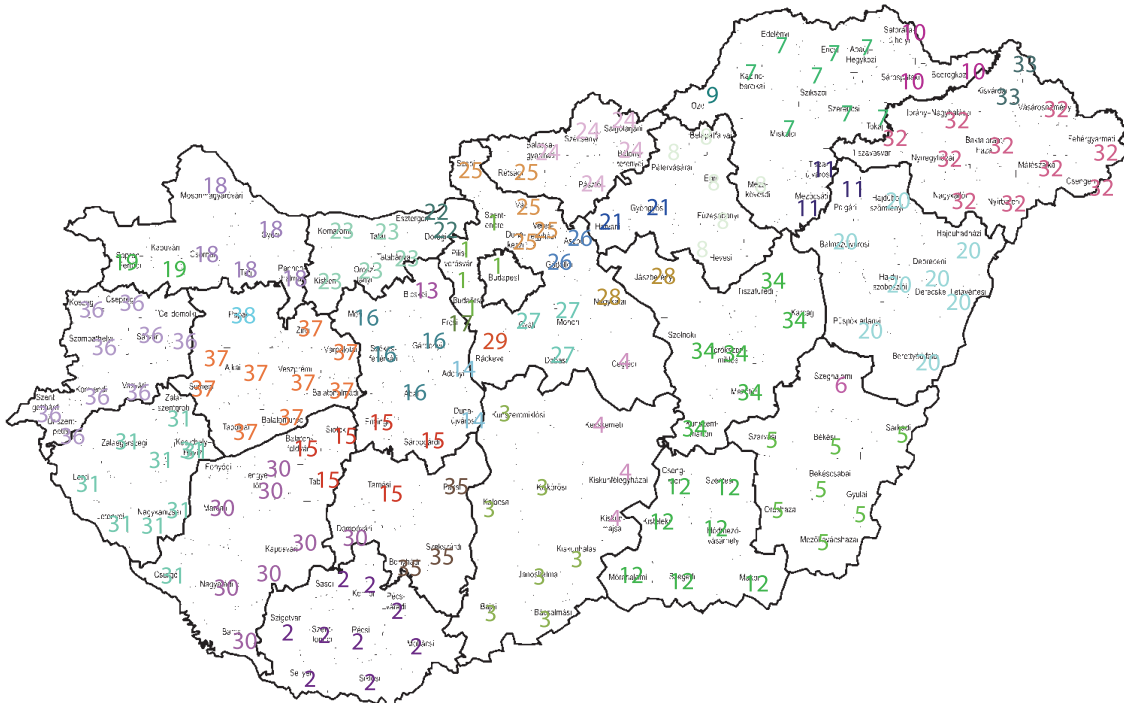
(b) Initial nodes are subregions ($l = 2$)

FIGURE 9: Communities formed by the Louvain method and Newman-Girvan (NG) null model ($I_i = k_i^{out}$ and $I_j = k_j^{in}$) reflect the settlement hierarchy as the resultant communities are almost identical to the counties.

In this case, the resultant communities will reflect another unmodelled surplus in the number of connections. When the attractiveness and the distances are considered in the null model, the communities will reflect the additional economic attractiveness/similarity of the regions.

As Figure 10 shows, the algorithm generates a huge cluster of a well developed regions with Budapest, the larger cities and county seats with high TDIs, and several small communities related to isolated and less developed subregions.
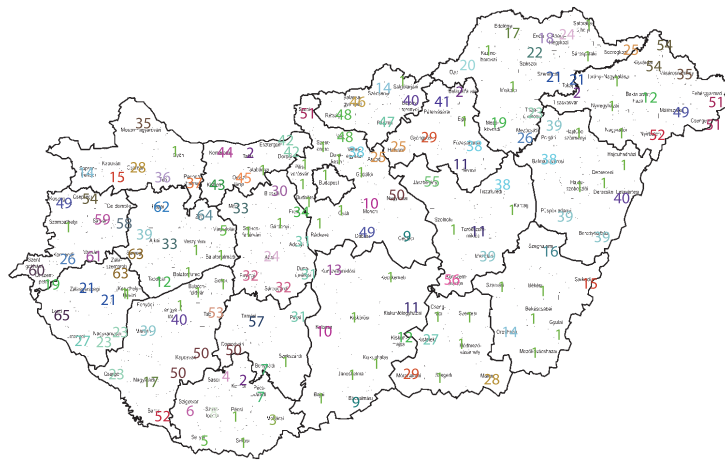
(a) TDI-based gravitational model: Initial nodes are subregions ($l = 2$)
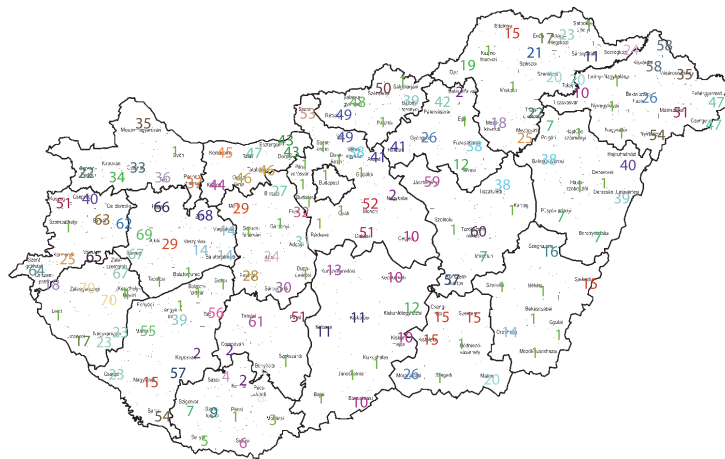
(b) The same TDI-based gravitational model at higher resolution $\gamma_r = 1.1$
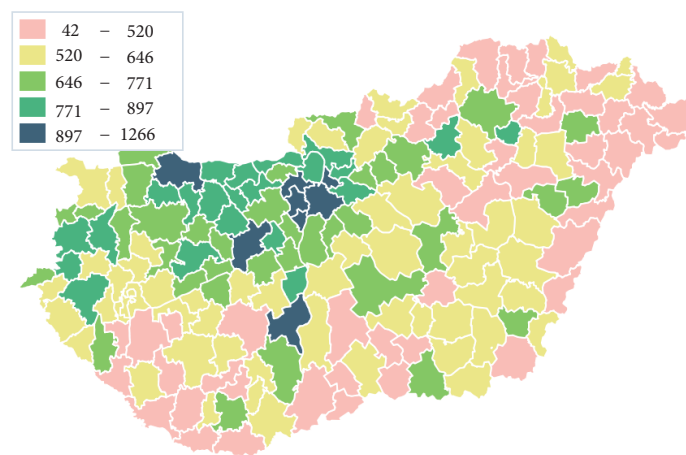
(c) Spatial distribution of the TDI per capita (in 1000 HUF)

FIGURE 10: Communities formed by the Louvain method and gravitational null models reflect the attractiveness of Budapest as only less developed closed regions were not assigned to the module of the capital.

4. Conclusions

Regional policy-making and monitoring are firm-centered, incentive-based, and state-driven. Personal investments define ties between geographical locations. We analyzed the structure of this ownership network and proposed a methodology to characterize regional attractiveness based on a set of null models identified to approximate the probabilities of link formation. According to the levels of the settlement hierarchy, a four-level study was conducted.

Based on the calculation of the internal and external network densities, several measures were proposed to evaluate the attractiveness and development of towns and geographical regions. The results indicate that small and less competitive regions have less internal connections, while larger cities are much more open.

To provide a more in-depth insight into the network, the dependence of link formation on distance was studied. The probability of connections between owners and their companies shows a much more rapid degree of distance decay than experienced in social networks. The attractiveness of the capital is so high that its connections are much less dependent on distance than other cities.

Based on the combination of three deterrence models and five sets of indicators, 15 different null models were identified besides the classical Newman-Girvan random configuration model. Communities statistically have more significant edge weights that would be wired according to the null model. As it was highlighted that underestimated link probabilities are the sources of modularity, a scatter plot was designed to visualize how the null model approximates the real structure of the network.

The identification of gravity-type models highlighted that link formation is nonlinearly dependent on the studied variables. Furthermore, the target nodes are much more important when determining the probability of link formation than the source nodes which also confirms why the structural analysis of company-ownership networks can be used to measure regional attractiveness.

We applied the Louvain community detection algorithm to form clusters of cities and subregions and compared the resultant communities to administrative regions. When the null model more closely approximates the real structure of the network, then the modularity is expected to be lower. As community detection forms modules whose internal link densities are significantly higher than what would be expected from the applied null models, spatial clusters that were highlighted by the distance independent random configuration model are almost identical to the counties. Communities generated based on the gravitational models, which correctly estimate the number of internal nodes and the dependence of link formation on distance, exploited the attractiveness of the capital, as they form a massive cluster that includes most of the centers of each county, bigger cities, and the competitive touristic regions, while the remaining small clusters reflect isolated regions that are less developed and less attractive.

Appendix

A. Internal and External Connection-Based Evaluation

Finding community structure means the assignment of the nodes into groups, where within the nodes are highly connected and across the nodes of the communities they are much loosely connected to each other [51].

The density of the whole network can be calculated as

$$D = \frac{L}{N^{[p]} N^{[co]}}. \tag{A.1}$$

while the internal density of the region is calculated as

$$D_i^{[l,in]} = \frac{a_{i,i}^{[l]}}{n_i^{[l,p]} n_i^{[l,co]}}. \tag{A.2}$$

$D_i^{[l,in]}/D$ compares internal complexity of the regions to the whole network.

The probability of an external tie, in other words, the external density, can be calculated in a similar fashion:

$$D_i^{[l,ex]} = \frac{\sum_{i \neq j} a_{i,j}^{[l]}}{N^{[l,p]} \left(N^{[l,co]} - n_i^{[l,co]} \right)}, \tag{A.3}$$

where $N^{[l,co]} - n_i^{[l,co]}$ represents the number of companies that are outside of the i-th region at the $[l]$-th level of the settlement hierarchy.

To evaluate the openness as a measure of the attractiveness of the region, the ratio of the external to internal probabilities can be defined as

$$O_i^{[l]} = \frac{D_i^{[l,ex]}}{D_i^{[l,in]}}. \tag{A.4}$$

Apart from taking into account internal and external links, the direction of the connections can be considered. Expansion computes the number of edges pointing outside the community [37]:

$$E_i^{[l]} = \frac{\sum_i a_{i,j}^{[l]} - \sum_i a_{i,i}^{[l]}}{n_i^{[l,p]}}. \tag{A.5}$$

Similarly, the ability of a community to collects links can be determined by the normalized number of links that point inside the community:

$$LCA_i^{[l]} = \frac{\sum_j a_{i,j}^{[l]} - a_{i,i}^{[l]}}{n_i^{[l,co]}}. \tag{A.6}$$

Cut ratio is similar to the internal density as it computes the fraction of edges pointing out and the number of possible edges that are pointing outside the community:

$$CR_i^{[l]} = \frac{\sum_j a_{i,j}^{[l]} - a_{i,i}^{[l]}}{n_i^{[l,p]} \left(N^{[l,co]} - n_i^{[l,co]} \right)}. \tag{A.7}$$

TABLE 4: Parameters of the power-law distributions fitted to networks at different settlement hierarchy levels.

Distribution	k_{sat}	k_{cut}	γ
$k_j^{[1,in]}$ (LAU 2)	120	15061	1.85
$k_j^{[1,out]}$ (LAU 2)	138	19709	1.87
$k_j^{[2,in]}$ (LAU 1)	1974	392724	2.04
$k_j^{[2,out]}$ (LAU 1)	2070	348339	2.04
$k_j^{[3,in]}$ (NUTS 3)	19693	392724	2.54
$k_j^{[3,out]}$ (NUTS 3)	20401	348339	2.49
$k_j^{[4,in]}$ (NUTS 2)	74161	557112	3.31
$k_j^{[4,out]}$ (NUTS 2)	77042	519967	3.35

B. Improvement of the Resolution

The modularity always increases when small communities are assigned to one group [52]. Modularity optimization with the null model p^{NG} has a resolution threshold which means it fails to identify small communities in large networks and communities consisting of less than ($\sqrt{L/2}$-1) internal links [53]. Reichardt and Bornholdt (RB) generalized the modularity function by introducing an adjustable γ_r parameter [54, 55] to handle this problem, which for our directed and weighted networks is

$$M_{RB}^{dir} = \frac{1}{L} \sum_i \sum_j \left(a_{ij} - \gamma_r \frac{k_i^{out} k_j^{in}}{L} \right) \delta \left(C_i, C_j \right). \quad (B.1)$$

Arenas, Fernandez, and Gomez (AFG) also proposed a multiresolution method by adding r self-loops to each node [56]. This algorithm increases the strength of a node without altering the topological characteristics of the original network, as $\mathbf{A}_r = \mathbf{A} + r\ \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix and r the weight of the self-loops of each node:

$$M_{AFG} = \frac{1}{L'} \sum_i \sum_j \left(a'_{i,j} - \frac{k_i^{out\prime} k_j^{in\prime}}{L'} \right) \delta \left(C_i, C_j \right), \quad (B.2)$$

where $L' = L + Nr$, $L = \sum_{i,j} a_{i,j}$, $k_i^{out\prime} = k_i^{out} + r$, $k_j^{in\prime} = k_j^{in} + r$, and

$$a'_{i,j} = \begin{cases} a_{i,j}, & \text{if } i \neq j, \\ a_{i,j} + r, & \text{if } i = j. \end{cases} \quad (B.3)$$

These methods still have the intrinsic limitation, so large communities may have been split before small communities became visible. The theoretical results indicated that this limitation depends on the degree of interconnectedness of small communities and the difference between the sizes of the communities, while being independent of the size of the whole network [52].

It should be noted that the modularity decreases when $p_{i,j}$ more closely approximates the real $a_{i,j}$ values which is equivalent to finding the null model that most closely fits.

C. Network Topology Analysis

The degree distribution was determined in all levels of the settlement hierarchy by following the methodology presented
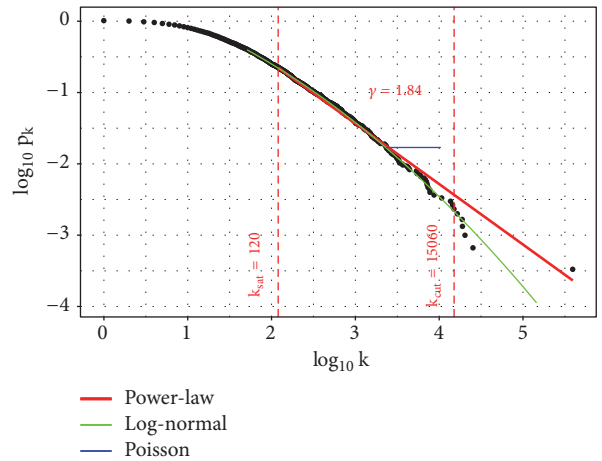
FIGURE 11: Distribution of the $k_j^{[1,in]}$ edges at the LAU 2 settlement hierarchy level.

in [13]. Figure 11 shows that the distribution shows small-degree saturation and high-degree cutoff. Several distribution functions were fitted. The two-sided Voung's test statistic [57] showed that exponential and Poisson distributions which reflect the randomness of connections could be rejected. According to this test, the power-law distribution cannot be rejected. The estimated parameters are shown in Table 4. The power-law distribution of the incoming and outgoing connections reflects the preferential attachment-type structure of the network.

In hierarchical networks, nodes with high degree tend to connect to nodes that are less connected to others [58]. Therefore, the hierarchical structure of the network is reflected by the dependence of the local clustering coefficient $C(k)$ on the degree of the nodes. As Figure 12 shows, $C(k)$ decreases with increasing k with $C(k) \approx k^{-0.3}$ which indicates the hierarchical structure of the network [58, 59].

D. Notations

p: Person/investor who is equivalent to the owner of a company

co: Company

[l]: Level of the settlement hierarchy (see (2))
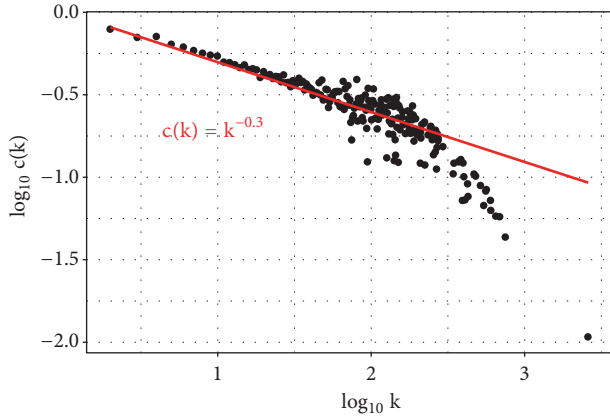
FIGURE 12: Local clustering coefficient as a function of the $k_j^{[1,in]}$ node degrees.

$entity^{[l]}$: Aggregation of an *entity* at level l of the settlement hierarchy

$\mathbf{A}^{[p,co]}$: Biadjacency matrix of person-company ownership network

$a_{i,j}^{[p,co]}$: An element (edge weight) of the $\mathbf{A}^{[p,co]}$ biadjacency matrix of person-company ownership network

$\mathbf{A}^{[p,l]}, \mathbf{A}^{[co,l]}$: Incidence matrices of person-location and company-location bipartite networks at the level l of the settlement hierarchy

$\mathbf{A}^{[l]}$: Simpler notation of an adjacency matrix of location network at l level of settlement hierarchy (see (3))

$k_j^{[l,in]}$: In-degree of the j-th node (geographic region) at level l of the settlement hierarchy

$k_i^{[l,out]}$: Out-degree of the i-th node (geographic region) at level l of the settlement hierarchy

$n_j^{[l,co]}, n_j^{[l,p]}$: Numbers of companies and people in the j-th region at level l of the settlement hierarchy

$N^{[co]}, N^{[p]}$: Number of companies and people/owners/investors in the network

L: Number of links in the network

C: Set of communities (each node is a member of exactly one community)

$C^{[l]}$: Set of communities at level l of the settlement hierarchy (C^1 denotes the set of towns)

$n_c^{[l]}$: Number of communities at level l of the settlement hierarchy

$f(C)$: Generally a metric as a function of community structure that indicates the goodness-of-fit of the community on the basis of the connectivity of nodes in it

$f(C^{[l]})$: Metric of the goodness-of-fit of the community structure which is the level l of the settlement hierarchy

M: A special $f(C)$ defined by (11) called modularity of network

M_c: Modularity of community c (sum of the modularity of each community yields the modularity M of the network)

$D_i^{[l,in]}, D_i^{[l,ex]}$: Internal and external densities of the i-th community at level l of the settlement hierarchy, defined by (A.2) and (A.3)

$O_i^{[l]}$: Openness of the i-th community at level l of the settlement hierarchy, defined by (A.4)

$E_i^{[l]}$: Expansion of the i-th community at level l of the settlement hierarchy, defined by (A.5)

$LCA_i^{[l]}$: Link-collection ability of i-th community at level l of the settlement hierarchy, defined by (A.6)

$CR_i^{[l]}$: Cut ratio of the i-th community at level l of the settlement hierarchy, defined by (A.7).

Data Availability

The data used to support the findings of this study are available from the website of the corresponding author (https://www.abonyilab.com/network-science/structural-analysis).

Disclosure

Conflicts of Interest

The authors declare that no conflicts of interest exist with regard to the publication of this paper.

Acknowledgments

References

[1] L. Kendrick, K. Musial, and B. Gabrys, "Change point detection in social networks—critical review with experiments," *Computer Science Review*, vol. 29, pp. 1–13, 2018.

[2] K. Musial, M. Budka, and K. Juszczyszyn, "Creation and growth of online social network: How do social networks evolve?" *World Wide Web*, vol. 16, no. 4, pp. 421–447, 2013.

[3] P. Bródka, K. Musial, and P. Kazienko, "A Method for Group Extraction in Complex Social Networks," in *Knowledge Management, Information Systems, E-Learning, and Sustainability Research*, vol. 111 of *Communications in Computer and Information Science*, pp. 238–247, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[4] M. Qin, D. Jin, D. He, B. Gabrys, and K. Musial, "Adaptive community detection incorporating topology and content in social networks," in *Proceedings of the 9th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2017*, pp. 675–682, Australia, August 2017.

[5] P. Kazienko, K. Musial, E. Kukla, T. Kajdanowicz, and P. Bródka, "Multidimensional Social Network: Model and Analysis," in *Computational Collective Intelligence. Technologies and Applications*, vol. 6922 of *Lecture Notes in Computer Science*, pp. 378–387, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[6] S. T. Cavusgil, T. Kiyak, and S. Yeniyurt, "Complementary approaches to preliminary foreign market opportunity assessment: Country clustering and country ranking," *Industrial Marketing Management*, vol. 33, no. 7, pp. 607–617, 2004.

[7] C. del Campo, C. M. F. Monteiro, and J. O. Soares, "The European regional policy and the socio-economic diversity of European regions: A multivariate analysis," *European Journal of Operational Research*, vol. 187, no. 2, pp. 600–612, 2008.

[8] A. Amin, "An Institutionalist Perspective on Regional Economic Development," *International Journal of Urban and Regional Research*, vol. 23, no. 2, pp. 365–378, 1999.

[9] M. Wang, "Location Is (Still) Everything: The Surprising Influence of the Real World on How We Search, Shop, and Sell in the Virtual One by David R. Bell," *Southeastern Geographer*, vol. 56, no. 4, pp. 476-477, 2016.

[10] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[11] G. Bianconi and A.-L. Barabási, "Competition and multiscaling in evolving networks," *EPL (Europhysics Letters)*, vol. 54, no. 4, pp. 436–442, 2001.

[12] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Structure of growing networks with preferential linking," *Physical Review Letters*, vol. 85, no. 21, pp. 4633–4636, 2000.

[13] A.-L. Barabßsi, *Network science book*, Center for Complex Network, Northeastern University, Boston, 2014, http://barabasi.com/networksciencebook.

[14] A. Blank and S. Solomon, "Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components)," *Physica A: Statistical Mechanics and its Applications*, vol. 287, no. 1-2, pp. 279–288, 2000.

[15] G. Duranton and D. Puga, "The Growth of Cities," *Handbook of Economic Growth*, vol. 2, pp. 781–853, 2014.

[16] X. Gabaix, "Zipf's law for cities: an explanation," *The Quarterly Journal of Economics*, vol. 114, no. 3, pp. 739–767, 1999.

[17] M. Cristelli, M. Batty, and L. Pietronero, "There is more than a power law in Zipf," *Scientific Reports*, vol. 2, article no. 812, 2012.

[18] M. Reba, F. Reitsma, and K. C. Seto, "Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000," *Scientific Data*, vol. 3, 2016.

[19] P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte, "Uncovering space-independent communities in spatial networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 108, no. 19, pp. 7663–7668, 2011.

[20] M. Barthélemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1–3, pp. 1–101, 2011.

[21] G. K. Zipf, *Human behavior and the principle of least effort*, Addison-Wesley Press, 1949.

[22] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora, "Distance matters: Geo-social metrics for online social networks," in *Proceedings of the 3rd Conference on Online Social Networks, WOSN'10*, USENIX Association, 2010.

[23] R. Lambiotte, V. D. Blondel, C. de Kerchove et al., "Geographical dispersal of mobile communication networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 21, pp. 5317–5325, 2008.

[24] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins, "Geographic routing in social networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 102, no. 33, pp. 11623–11628, 2005.

[25] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: improving geographical prediction with social and spatial proximity," in *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, pp. 61–70, ACM, Raleigh, NC, USA, April 2010.

[26] J. Goldenberg and M. Levy, *Distance is not dead: Social interaction and geographical distance in the internet era*, 2009, https://arxiv.org/abs/0906.3202.

[27] A. Reynolds-Feighan and P. McLay, "Accessibility and attractiveness of European airports: A simple small community perspective," *Journal of Air Transport Management*, vol. 12, no. 6, pp. 313–323, 2006.

[28] A. P. Groh and M. Wich, "A Composite Measure to Determine a Host Country's Attractiveness for Foreign Direct Investment," *SSRN Electronic Journal*.

[29] C. E. Gearing, W. W. Swart, and T. Var, "Establishing a Measure of Touristic Attractiveness," *Journal of Travel Research*, vol. 12, no. 4, pp. 1–8, 1974.

[30] J. E. Anderson, "The gravity model," *Annual Review of Economics*, vol. 3, pp. 133–160, 2011.

[31] J. E. Anderson, "A theoretical foundation for the gravity equation," *American Economic Review*, vol. 69, no. 1, pp. 106–116, 1979.

[32] K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski, and S. S. Manna, "The international trade network: Weighted network analysis and modelling," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 2, 2008.

[33] M. Levy, "Scale-free human migration and the geography of social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 21, pp. 4913–4917, 2010.

[34] W.-S. Jung, F. Wang, and H. E. Stanley, "Gravity model in the Korean highway," *EPL (Europhysics Letters)*, vol. 81, no. 4, Article ID 48005, 6 pages, 2008.

[35] A. Wahid -Ul- Ashraf, M. Budka, and K. Musial-Gabrys, "Newton's Gravitational Law for Link Prediction in Social Networks," in *Complex Networks & Their Applications VI*, vol. 689 of *Studies in Computational Intelligence*, pp. 93–104, Springer International Publishing, Cham, 2018.

[36] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.

[37] T. Chakraborty, A. Dalmia, A. Mukherjee, and N. Ganguly, "Metrics for Community Analysis," *ACM Computing Surveys*, vol. 50, no. 4, pp. 1–37, 2017.

[38] E. A. Leicht and M. E. J. Newman, "Community structure in directed networks," *Physical Review Letters*, vol. 100, no. 11, Article ID 118703, 2008.

[39] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 3, Article ID 036104, 19 pages, 2006.

[40] R. Cazabet, P. Borgnat, and P. Jensen, "Enhancing Space-Aware Community Detection Using Degree Constrained Spatial Null Model," in *Complex Networks VIII*, Springer Proceedings in Complexity, pp. 47–55, Springer International Publishing, Cham, 2017.

[41] X. Liu, T. Murata, and K. Wakita, *Extending modularity by incorporating distance functions in the null model*, 2012, CoRR, abs/1210.4007.

[42] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Article ID P10008, 2008.

[43] P. Schuetz and A. Caflisch, "Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, no. 4, Article ID 046112, 2008.

[44] B. Lengyel, A. Varga, B. Ságvári, Á. Jakobi, and J. Kertész, "Geographies of an online social network," *PLoS ONE*, vol. 10, no. 9, 2015.

[45] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowledge and Information Systems*, vol. 42, no. 1, pp. 181–213, 2015.

[46] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel, "Urban gravity: A model for inter-city telecommunication flows," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 7, 2009.

[47] D. Balcan, V. Colizza, B. Gonçalves, H. Hud, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 106, no. 51, pp. 21484–21489, 2009.

[48] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, "The complex network of global cargo ship movements," *Journal of the Royal Society Interface*, vol. 7, no. 48, pp. 1093–1103, 2010.

[49] I. S. Jutla, L. G. Jeub, and P. J. Mucha, *A generalized Louvain method for community detection implemented in MATLAB*, 2011, http://netwiki.amath.unc.edu/GenLouvain/GenLouvain.

[50] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, no. 5980, pp. 876–878, 2010.

[51] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.

[52] J. Xiang and K. Hu, "Limitation of multi-resolution methods in community detection," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 20, pp. 4995–5003, 2012.

[53] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 1, pp. 36–41, 2006.

[54] J. Reichardt and S. Bornholdt, "Detecting fuzzy community structures in complex networks with a potts model," *Physical Review Letters*, vol. 93, no. 21, 2004.

[55] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 1, Article ID 016110, 2006.

[56] A. Arenas, A. Fernández, and S. Gómez, "Analysis of the structure of complex networks at different resolution levels," *New Journal of Physics*, vol. 10, Article ID 053039, 2008.

[57] Q. H. Vuong, "Likelihood ratio tests for model selection and nonnested hypotheses," *Econometrica*, vol. 57, no. 2, pp. 307–333, 1989.

[58] E. Ravasz and A. Barabási, "Hierarchical organization in complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 67, no. 2, Article ID 026112, 2003.

[59] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Pseudofractal scale-free web," *Physical Review E*, vol. 65, no. 6, pp. 66–122, 2002.

Research Article

Crisis Spreading Model of the Shareholding Networks of Listed Companies and Their Main Holders and Their Controllability

Yuanyuan Ma [1] **and Lingxuan Li** [2]

[1]*School of Economics, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China*
[2]*School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China*

Correspondence should be addressed to Yuanyuan Ma; mayuanyuan@mail.neu.edu.cn

Bankruptcy of listed companies or shareholders delisting usually causes the crisis spreading in stock markets. Based on the systematic analysis of the epidemic diseases and rumors spreading on the complex networks, the SIR model is introduced to research the crisis spreading in shareholding networks of listed companies and their main holders on the basis of the data about ownership structure in Chinese Stock Markets. The characteristics of shareholding networks are studied, and the parameters for the SIR model are obtained by empirical approach. Then, the numerical computation method is successfully used to analyze the crisis spreading in the networks when the networks meet random failures or intentional attacks. We find the networks have good robustness against the random failures. However, the crisis will spread at a high speed and cause catastrophic damage if there are some failures or attacks on hub vertices in the networks. Under this condition, the networks show obvious vulnerability. Last but not least, the controllability of the networks under the condition of intentional attacks and random failures is studied. The results show that if the network is controlled globally, it is more reliable to allow a politically good new or an appropriate exciting economical policy to play the role in orienting markets under the control of public opinions as the crisis occurs. However, under normal circumstances, controlling a small part of driver vertices representing listed companies, applying appropriate control strategies, and using its characteristics of high efficiency of sending information can effectively control the stock market. Our research provides a new reference to further exploration about the transmission mechanism of the crisis based SIR model and further research on the controllability of crisis spreading in financial markets.

1. Introduction

Reviewing the long course of human history, each financial crisis has led to economic disaster. The Great Depression which started about 1929 and lasted until the late 1930s or early 1940s swept through all the countries in Western Europe and the United States. When the members of Organization of Arab Petroleum Exporting Countries proclaimed an oil embargo against the United States along with the fourth Arab-Israeli War breaking out, the 1973 oil crisis started, which subsequently led to economic crisis. The Latin American debt crisis occurred in the early 1980s (and for some countries starting in the 1970s) known as the "lost decade". Around 1990s, the Japanese asset price bubble collapsed because of the great inflation of real estate and stock prices.

The Asian financial crisis gripped most area in Asia from 1997 to 1998.

With Chinese joining in the WTO, our financial market opened to the outside world further. The relationship between Chinese mainland financial system and foreign financial systems has been getting closer. The Chinese stock markets can be influenced by various kinds of crises from aboard. For example, the US subprime mortgage crisis in 2007 triggered the worldwide financial crisis. The closing price of Shanghai securities composite index (index code: 000001) falls from the maximum point 6082.06 CNY/point on October 16, 2007 to 1706.7 CNY/point on November 4, 2008. The closing price of the Shenzhen composite index (index code: 399106) falls from the maximum point 1576.5 CNY/point On January 15, 2008 to 456.97 CNY/point on November 4, 2008. The

samples of the index 000001 and index 399106 include all the issued shares in Shanghai security exchange and Shenzhen security exchange, respectively. It means that more than 70% Market Capitalization had vaporized during the year of 2008.

Invariably, each financial crisis will do very serious damage to the country's real economy. If we can completely master the transmission mechanism of financial crisis and exactly predict the financial crisis, it is possible for government to take steps to nip the crisis in the bud.

It is well known that the stock markets are the barometer of national economic development. The stock prices of listed companies can reflect the capital demand and supply situation, market demand, current situation, anticipation of industry development trend, and the unrest of political situations. The stock markets are so sensitive to the economic crisis and financial crisis that once some abnormal phenomena occur in the stock market, the real economy of the country will be affected inevitably and seriously. It is easy for the abnormality in stock market to trigger a global financial crisis or economic crisis. On the contrary, if some abnormal phenomena occur in the real economy, the bubble in the stock markets will collapse firstly. Then the crisis will penetrate into every aspect of people's lives rapidly. Therefore, studying the structural characteristics of stock markets and the crisis spreading in the stock markets are important.

Along with the study of the transmission mechanism and the statistical mechanics of complex networks, two far-reaching spreading models are formed. They are susceptible-infected-susceptible (SIS) mode [1–4] and SIR model [1, 5–7]. When talking about SIS and SIR model on networks, the literature typically refers to epidemiological dynamical processes which have been studied for quite a long time. As for the so-called SIR model, it means that each vertex in networks lies in one of the following states: susceptible (healthy state), infected, removed (or refractory, or recovered). At each step, the susceptible vertices become infected vertices with certain probability if they enter in contact with infective vertices and disease transmission occurs successfully. At the same time, the infected vertices become removed (or refractory, or recovered) vertices with certain probability [1, 5–7]. Later, the SIR and SIS model are also introduced to describe the rumor spreading process in interpersonal networks since they have the similar spreading process of epidemiology.

Bankruptcy of listed companies or shareholders delisting usually causes the crisis spreading in stock markets. Based on the systematic analysis of the epidemic diseases and rumor spreading on the complex networks, the SIR model will be introduced to research the crisis spreading in shareholding networks of listed companies and their main holders.

Compared with the SIR model of epidemic spreading or rumor spreading on general complex networks, the susceptible-infected-removed (SIR) model of crisis spreading shows great differences in the shareholding networks due to stock markets having their own characteristics. The networks are established by the real data of the mutual investment relationships between the listed companies and their main holders, and the shareholding networks are the typically weighted and directed networks. In order to reduce the loss, the vertices representing the listed companies or the main holders make different decisions according to their different situations. Meanwhile, the decisions made by the listed companies or the main holders may be obviously different. Therefore, the susceptible or infected vertices in the correlated networks will become infected vertices or removed vertices with no certain given probability, respectively. Not only is the SIR model beneficial to master the transmission mechanism of crisis spreading on the stock markets, but also it can help to reveal the dissemination process and root cause of financial crisis.

Compared to model research of crisis spreading, perhaps reflecting the controllability and control strategies for the crisis are more concerned. The ultimate goal of studying complex network systems is controlling, or manual intervention. It involves issues such as controllability, control strategies, precise control, minimum cost control, and spontaneous controllability. In 2011, Liu and the control theory community J. J Slotine and the complex network leader A. L. Barabasi cooperated to use the linear system structure controllability theory and introduced mapping maximum matching for networks and Kalman's controllability rank condition to establish the theory for analyzing the controllability of complex networks [8]. On the basis of Liu's work, Jia et al. divide the network vertices into three categories, critical vertices, intermittent vertices, and redundant vertices, and further calculate the proportion of the three types of driver vertices and propose the concept of control capacity [9, 10]. Subsequently, Yuan et al. propose a more accurate concept of networks controllability based on the PBH rank criterion and further introduced it into the research of multirelational networks and multilayer networks [11]. Sun et al. consider that, in the control design process of practical complex systems, the control problem of the system is usually considered based on the energy optimal control, which involves the calculation of the controllable matrix, and thus controllability theory based on the singularity of Gramian matrix is proposed [12]. Based on these three types of controllability research, many researchers have carried out various kinds of research.

In this paper, we will not only carry on model research of crisis spreading by establishing a SIR model of crisis spreading in stock markets, but also research the controllability and controlling strategies of shareholding networks when the networks meet random failures or intentional attacks.

This paper can be divided into 7 sections. Section 1 is the introduction about the related subjects and current research. The data and the methods of establishing networks in this paper are elaborated in Section 2. In Section 3, based on the analysis of crisis spreading in the shareholding networks, the SIR model in stock markets is established. Section 4 introduces the characteristics of shareholding networks and the parameters selection for the SIR model. Based on SIR model, Section 5 simulates the process of crisis spreading when the networks meet random failures and intentional attacks. And its controllability of networks has been studied in Section 6. The main conclusions and some related discussions are given in Sections 7 and 8.

2. Data and Networks

2.1. Data. The data are extracted from the RESSET Financial Research Database (http://www.resset.cn/), including all the stocks issued in the Shanghai Stock Exchange and the Shenzhen Stock Exchange before December 31, 2009. The documents we selected include the Main Stockholders List and Ownership Structure and the Yearly Market Capitalization (CNY) of all listed companies.

Through the issuance of stocks, the shareholding networks between the listed companies and their shareholders are formed. The shareholders of a listed company may be other listed companies, funds, non-listed enterprises, individuals, universities, etc. For each stock, the number of the shareholders in stock markets is numerous. The shareholders holding less proportion stocks may buy or sell their stocks frequently in each trading day. Thus, the shareholders' information constantly changes during exchange hours. To handle all the information of the listed companies and their holders is incredible. However, it is worth noting that the proportion of stocks owned by the major holders of each listed company is more than 55%. These data can be accessed from the Internet or related financial database. According to our statistics, the proportion of average shares of the 10th largest holders of the listed companies is smaller than 0.42%. Comparatively, the holders holding share smaller than 0.42% have little influence on the analysis of the community structures and topological characteristics of complex networks.

There is no uniformity in naming convention about the holders' name for the annual reports in the Securities and Futures Commission (SFC), such as full name or abbreviated name, name in Chinese or in English, subsidiary company's name or parent company's name, etc. Therefore, the name of the same holder must be unified. The names of all domestic companies are unified in Chinese. The names of all foreign companies are unified in English. If some different holders are the subsidiary companies of a certain company, we view them as the same vertex in the networks. For example, China Life Insurance Company Limited and China Life Asset Management Company Limited are the subsidiary companies of China Life Insurance (Group) Company. Under this condition, the three companies should be viewed as the same vertex representing their parent company (China Life Insurance (Group) Company).

2.2. Networks. In shareholding networks, the listed companies and their main holders are the vertices, which is different to [13]. The shareholding relationships are the edges of the networks. According to the graph theory [14], the networks can be indicated by the directed graph $G = (V, E)$, where V is a set of vertices, which represents the listed companies and their main holders. E is the ordered pair of vertices, called directed edges, which represents investment relationships.

For the purpose of indicating the mutual investment relationships in stock markets, the basic subgraph of shareholding networks is extracted, as shown in Figure 1.

In Figure 1, the symbol of hexagons denotes a listed company; the symbol of circles denotes a main holder except
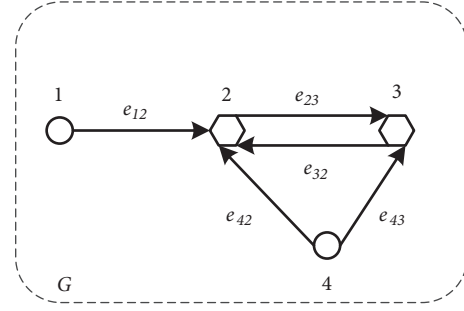
Figure 1: A subgraph of shareholding networks.

the listed companies (in this paper, denoted as nonlisted holders). The set V includes two kinds of vertices: the set of listed companies (denoted by V_L) and the set of main holders (denoted by V_S), where $V_L \subset V, V_S \subseteq V$. It is worth noting that some listed companies can also act as the main holders of some other listed companies, such as Company 2 and Company 3 as shown in Figure 1. Thus, we have $v_2, v_3 \in V_L$ and $v_1, v_2, v_3, v_4 \in V_S$.

The set E indicates the mutual investment relationships between the listed companies and their main holders. If holder i holds certain proportion of stocks issued by the listed company j, holder i has invested in the listed company j. In the shareholding networks, the investment relationships can be represented by the ordered pair $e_{ij} = (v_i, v_j)$; the direction is from vertex v_i to vertex v_j. Meanwhile, the weight between vertex v_i and vertex v_j is denoted as the symbol u_{ij}. u_{ij} is the proportion of the holding shares of holder i to the total issued shares by the listed company j. $u_{ij} = 0$ indicates that there are no investment relationships between the holder i and the listed company j. If $e_{23} \neq 0$ and $e_{32} \neq 0$, $e_{23} = (v_2, v_3)$ and $e_{32} = (v_3, v_2)$ will represent the difference investment relationships in the shareholding networks. In order to indicate the possible existing investment relationships in the shareholding networks, all kinds of shareholding relationships are listed in Table 1 on the basis of analyzing Figure 1.

The Yearly Market Capitalization of any listed company is the price of stocks issued by the listed company multiplied by the total issued shares. Suppose M_j as the Yearly Market Capitalization of the listed company j. The asset of holder i investing to listed company j is u_{ij} multiplied by M_j. Now, we can define the in-degree assets of listed company j as investment of their holders. We use s_j^{in} to represent the in-degree assets of listed company j, then

$$s_j^{in} = \left(u_{1j} + u_{2j} + \cdots + u_{ij} + \cdots + u_{N_L j} \right) \times M_j$$
$$= \sum_{i=1}^{N_L} u_{ij} \times M_j \tag{1}$$

where N_L is the total number of the listed companies. By the way, the nonlisted holders have not the in-degree assets.

TABLE 1: The existent investment relationships in the shareholding networks.

Number	Investment relationships	Set of vertices	Set of edges
1	A non-listed holder investing to one listed company	$\{v_1, v_2\}$; $v_1 \in V_S$; $v_2 \in V_L$	$\{e_{12}\}$
2	A non-listed holders investing to more than one listed company	$\{v_2, v_3, v_4\}$; $v_2, v_3, v_4 \in V_S$; $v_2, v_3 \in V_L$	$\{e_{42}, e_{43}\}$
3	A listed company investing to other listed company	$\{v_2, v_3\}$; $v_2, v_3 \in V_S$; $v_2, v_3 \in V_L$	$\{e_{23}, e_{32}\}$

Let the out-degree of nonlisted holder (or listed company) i as its investment to the other listed companies; we use s_i^{out} to represent the nonlisted holder (or listed company) i; then

$$s_i^{out} = u_{i1} \times M_1 + u_{i2} \times M_2 + \cdots + u_{ij} \times M_j + \cdots$$
$$+ u_{iN_s} \times M_{N_L} = \sum_{j=1}^{N_L} \left(u_{ij} \times M_j \right) \tag{2}$$

3. Analysis of Crisis Spreading in Shareholding Networks and the SIR Model

3.1. Susceptible State, Infected State, and Removed State in Shareholding Networks. As mentioned above, SIR models are mainly used to investigate the epidemiological dynamical processes or the rumor spreading processes in social networks. The similar modeling method is introduced to research the crisis spreading processes on the basis of the networks of listed companies and their main holders. Thus, the definitions of the vertices and edges of the networks should be carefully illustrated. The vertices represent the listed companies and the main holders. The objective existent shareholding relationships are the edges of the networks. Thus, the susceptible state, infected state, and removed state in the shareholding networks can be defined as follows.

Susceptible State. For the vertex representing nonlisted holders, it means that the external investment of the vertex has not changed. In other words, the out-degree assets of vertex and the directed edges from vertex to other vertices have not changed. For the vertex representing listed companies, it means that the Market Capitalization of the vertex has not decreased.

Infected State. For the vertex representing nonlisted holders, it means that the out-degree assets of the vertex have decreased; meanwhile, the directed edges from vertex to other vertices may change. For the vertex representing listed companies, it means that the Market Capitalization of the vertex has decreased.

Removed State. For the vertex representing nonlisted holders or listed companies, it means that the vertices are removed from the networks.

It is noteworthy that the susceptible vertices may become infected vertices or removed vertices directly when the states of vertices have changed in the shareholding networks. For example, the holder i only holds the stocks issued by a certain listed company j. When the Market Capitalization of company j descend because of mismanagement in business, the state of the vertex v_i will become susceptible state if the listed company j is still allowed to be listed on the stock markets. However, the state of the vertices v_i and v_j will become the removed state if the company goes bankrupt.

In the next sections, each kind of the crisis spreading in shareholding networks is analyzed carefully, and the corresponding functions are obtained. To help the reader understand, this article will explain the corresponding change rules of vertices among susceptible state, infected state, and

removed state in detail on the basis of the typical network in Figure 1.

3.2. The Existent Failures in Networks. According to graph theory [14], the existent failures in networks can be divided into two categories.

3.2.1. Cut Edges. Suppose e is an edge in graph G, and a cut edge e of graph G means deleting the edge e from graph G; it can be denoted by $G - e$. If $T = \{e_1, e_2, \cdots, e_i,\}$ is an edge subset of E in graph $G = (V, E)$, deleting the edge subset T from subgraph G can be denoted by $G - T$.

3.2.2. Cut Vertices. A cut vertex v_i of graph G means deleting the vertex v_i together with the related incident edges, and the graph will be denoted by $G - v_i$. Correspondingly, a cut vertex set $C = \{v_1, v_2, \cdots, v_i\}$ of graph G means deleting the vertices $C = \{v_1, v_2, \cdots, v_i\}$ together with the related incident edges, and the subgraph will be denoted by $G - C$.

3.3. Crisis Spreading Model of SIR in Shareholding Networks. Compared to the SIR model of epidemic spreading or rumor spreading in complex networks, the SIR model of crisis spreading shows no given susceptible or infected probabilities in shareholding networks, in which the crisis spreading between any two vertices is influenced by many factors, such as shareholding proportion, Market Capitalization, in-degree assets, and out-degree assets. Meanwhile, the crisis spreading has obvious directions in the shareholding networks because the networks are established by the real data of the mutual investment relationships between the listed companies and their main holders.

After careful analysis, the failure in stock markets can be divided into 5 categories. Correspondingly, the crisis spreading functions in shareholding networks can be obtained as follows.

3.3.1. Influence of Nonlisted Holders' Bankruptcy. Suppose nonlisted holder i goes bankrupt; in shareholding networks the vertex v_i should be deleted; meanwhile, the edges linking vertex v_i to other vertices should also be deleted.

According to the proportion u_{ij} of the holding shares of holder i to the total issued shares by listed company j, the listed company j may be influenced. This kind of influence can be denoted by equation $C_2(v_i, v_j, u_{ij})$. Then, the bankruptcy of nonlisted holder i can be described as in

$$H_1 \left(v_i \right) = \begin{cases} G - v_i \\ C_2 \left(v_i, v_j, u_{ij} \right) \end{cases} \tag{3}$$

where $j = 1, 2, 3, \cdots, n$, and $v_j \in V_L$; $v_i \in V_S, v_j \notin V_L$. n is the total number of vertices before we delete the vertex v_i. If there is no investment relationship between the nonlisted holder i and listed company j, $u_{ij} = 0$. Under this condition, let $C_2(v_i, v_j, u_{ij}) = 0$. That is to say, the nonlisted holder i which goes bankrupt has no direct influence on listed company j.

In order to improve the replicability of this paper and master the functions and parameters for casual readers, five figures and many paragraphs are added to illustrate the rules
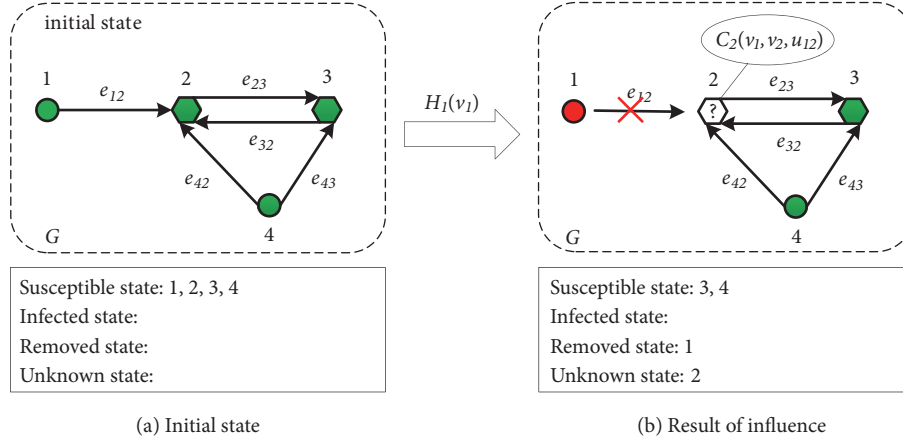
(a) Initial state   (b) Result of influence

FIGURE 2: Influence of bankruptcy of nonlisted holder 1 on the network.

of crisis spreading in Section 3.3 (Sections 3.3.1 to 3.3.5) based on simple network of Figure 1. The influence of the bankruptcy nonlisted holder 1 on the network is expressed in Figure 2.

Figure 2 shows that the bankruptcy of nonlisted holder 1 deletes the vertex v_1 together with the incident edge e_{12}; correspondingly, the susceptible state of vertex 1 becomes removed state directly; it also affects the investment relationship of company 2 linked to holder 1, which is denoted by $C_2(v_1, v_2, u_{12})$.

3.3.2. Bankruptcy of Listed Companies.
When listed company j goes bankrupt, the vertex v_j and the edges linking vertex v_j to other vertices will be deleted inevitably in the shareholding networks. At the same time, the vertices linked to vertex v_j will be influenced.

On the one hand, if listed company j holds the stocks issued by another listed company k, the influence of bankrupt company j on listed company k is $C_2(v_j, v_k, u_{jk})$, where $k = 1, 2, 3, \cdots, n, k \neq j$.

On the other hand, the total assets of the holders of listed company j will decrease inevitably. The loss assets of holder i, who has invested to listed company j, can be obtained as follows:

$$m_i = M_j \times u_{ij} \tag{4}$$

where M_j is the Market Capitalization of listed company j before going bankrupt.

At this time, holder i will make a decision for benefiting itself according to its situation of loss, which can be described by equation $H_2(v_i, m_i)$, where $i = 1, 2, 3, \cdots, n$ and $v_i \in V_S$.

As the analysis above, the changes of the shareholding networks can be described as (5) when the listed company j goes bankrupt.

$$
C_1(v_j)
$$
$$
= \begin{cases} G - v_j \\ H_2(v_i, m_i) & i = 1, 2, 3, \cdots, n \ \& \ i \neq j \\ C_2(v_j, v_k, u_{jk}) & j = 1, 2, 3, \cdots, n \ \& \ k \neq j \end{cases} \tag{5}
$$

where $v_i \in V_S$ and $v_j, v_k \in V_L$.

Using Figure 1 as an example, the influence of the bankruptcy of listed company 2 is shown in Figure 3.

The bankruptcy of listed company 2 makes the edges linking vertex v_2 to other vertices be deleted inevitably, including $e_{12}, e_{32}, e_{42}, e_{23}$, and the susceptible state of nonlisted holder 1 becomes a removed state because holder 1 only holds the stocks issued by listed company 2 and affects decision behavior of nonlisted holder 4 and listed company 3 because of their assets decreasing, marked by $H_2(v_2, m_4)$ and $C_2(v_2, v_3, u_{23})$, respectively.

3.3.3. Analysis of the Decision Behavior of Holders When Their Assets Decrease.
Under the following two conditions, the assets of holder i will decrease inevitably. ① Holder i holds more than one kind of stocks issued by different listed companies. The total assets of the holder will decrease inevitably when one of the listed companies goes bankruptcy. ② If holder i holds the stocks issued by listed company j, the assets of holder i will also be decreased inevitably when the Market Capitalization of listed company j descends because of mismanagement in business or some other reasons. Under these two conditions, the holder i will make a decision for benefiting itself, so that it can decrease the losses to a tolerant level.

Suppose m is the loss of assets of holder i before making the decision (m can be obtained from (6) and (10) for the two conditions mentioned above, respectively.). The holder i will make the decision according to the proportion of the loss to the total external investing assets of holder i (denoted as r_i). If M_j is the Market Capitalization of listed company j, r_i can be obtained from the following equation:

$$
r_i = \frac{m_i}{m_i + \sum_{\substack{j=1 \\ i \neq j}}^{n} \left(u_{ij} \times M_j \right)} \tag{6}
$$

where $v_i \in V_S$, $v_j \in V_L$, $j \neq i$. n is the total number of the listed companies. If holder i does not invest in the listed company j, $u_{ij} = 0$.

According the value of r_i, the loss of holder i can be divided into 3 kinds: general loss, heavy loss, and catastrophic

(a) Initial state

(b) Result of iinfluence
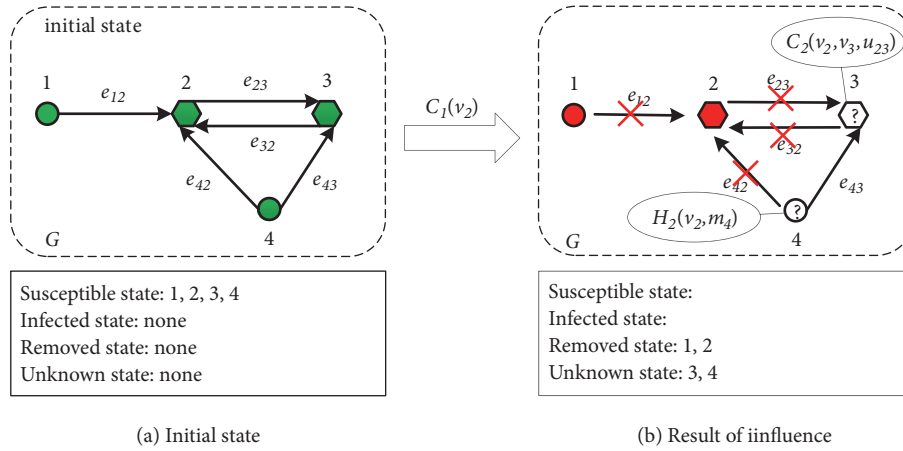
FIGURE 3: Influence of bankruptcy of listed company 2 on the network.

(a) Heavy loss of holder 4 (vertex 4)

(b) Catastrophic loss of holder 4 (vertex 4)
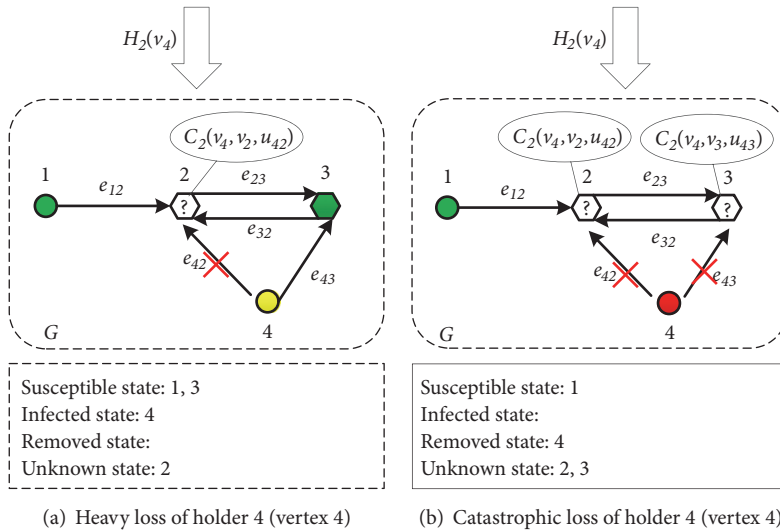
FIGURE 4: Decision behavior of holder 4 when it suffers a heavy or catastrophic loss.

loss. Correspondingly, the decision behavior of holder i can be described by the following equation:

$$
H_2(v_i, m_i)
$$
$$
= \begin{cases} 0 & 0 \le r_i < x_1 \\ G - e_{ij} \text{ and } C_2(v_i, v_j, u_{ij}) & x_1 \le r_i < x_2 \\ H_1(v_i) & r_i \ge x_2 \end{cases} \quad (7)
$$

where $i \ne j$, $v_i \in V_S$ and $v_j \in V_L$. x_1 and x_2 can be obtained from Section 4.2.

Equation (7) can be described as follows.

If $r_i < x_1$ (general loss), the holder i will make no new decision. Correspondingly, the vertex v_i will keep the same state as before in the shareholding networks.

If $x_1 \le r_i < x_2$ (heavy loss), holder i will sell one kind of stocks issued by company j. The assets of holder i investing in listed company j are minimum (not including zero)

compared to its investment to other companies. Correspondently, in the networks, the edge e_{ij} will be deleted and the influence of holder i selling the stocks issued by company j on company j can also be described by equation $C_2(v_i, v_j, u_{ij})$.

If $r_i \ge x_2$ (catastrophic loss), the holder i will sell all the holding stocks. It can be described by equation $H_1(v_i)$.

For the last two conditions, they can be mastered by the example in Figure 4.

The initial state of the networks is the same as Figure 3(a). Assuming the assets of holder 4 investing in company 2 are smaller than those in company 3, when holder 4 suffers a heavy loss, it will sell the stocks issued by company 2; correspondingly, the influence on company 2 expressed by $C_2(v_4, v_2, u_{42})$ and the susceptible state of vertex 4 becomes an infected state, as shown in Figure 4(a). When holder 4 gets a catastrophic loss, it will sell all the holding stocks and influence on the listed companies, as denoted by $C_2(v_4, v_2, u_{42})$ and $C_2(v_4, v_3, u_{43})$ in Figure 4(b).

3.3.4. Influence of Holder i on the Networks When Holder i Goes Bankrupt or Sells the Stocks Issued by Listed Company j. When holder i goes bankrupt or sells the stocks issued by listed company j, the Market Capitalization of listed company j may be influenced. Suppose M_j is the Market Capitalization before holder i goes bankrupt or sells the stocks issued by listed company j and M'_j is the new Market Capitalization after holder i goes bankrupt or sells the stocks issued by listed company j. Suppose R_j is the proportion of M'_j to M_j. Then,

$$R_j = \frac{M'_j}{M_j} \tag{8}$$

According to the Chinese Stock Exchange Listing Rules, the behaviors of selling or buying assets of the holders who hold the proportion of shares more than 5% of a certain listed company are viewed as major events in stock markets. The behaviors of the holder, who holds the proportion of shares more than 50% of a certain listed company, selling or buying assets are viewed as the behaviors of the listed company [15]. Therefore, we neglect the influence, which the proportion of single transaction shares of a certain listed company to its total shares is less than 5%. After careful analysis of the block trade of stocks in the stock markets, the new Market Capitalization M'_j is obtained, as shown in (12).

Equation (8) shows $R_j \geq 0$. Moreover, if R_j is more closer to 1, the fluctuation of the Market Capitalization of listed company j is smaller, so the damage to company j is smaller. When R_j is a certain value x_3, the listed company j will go bankrupt. The value of x_3 can be found in Section 4.2.

Therefore, when holder i goes bankrupt or sells the stocks issued by listed company j, the influence of holder i on listed company j can be written as in the following equation:

$$C_2\left(v_i, v_j, u_{ij}\right)$$
$$= \begin{cases} 0 & u_{ij} < 0.05 \\ C_1\left(v_j\right) & u_{ij} \geq 0.05 \text{ and } 0 \leq R_j \leq x_3 \\ C_3\left(v_j\right) \text{ and } H_2\left(v_q, m_q\right) & u_{ij} \geq 0.05 \text{ and } 1 > R_j > x_3 \\ 0 & R_j \geq 1 \end{cases} \tag{9}$$

where $q = 1, 2, 3, \cdots, n, q \neq j, i \neq j, i \neq q, v_i, v_q \in V_S, v_j \in V_L$.

The meanings of (9) are as follows.

If $u_{ij} < 0.05$, the behaviors of holder i going bankrupt or selling the stocks issued by company j will have no direct influence on company j.

If $u_{ij} \geq 0.05$ and $0 \leq R_j \leq x_3$, the behaviors of holder i going bankrupt or selling the stocks issued by company j will make company j go bankrupt or be delisted from the stock markets.

If $u_{ij} \geq 0.05$ and $1 > R_j > x_3$, the behaviors of holder i going bankrupt or selling the stocks issued by company j will make the Market Capitalization of company j decrease, but the company j will not go bankrupt. At this time, the company j will make different decisions according to its decrement of Market Capitalization (denoted as $C_3(v_j)$, where, $v_j \in V_L$). Meanwhile, the holders of company j will

make different decisions according to its assets decrement, which is the same as $H_2(v_q, m_q)$, where $q = 1, 2, 3, \cdots, n$, $q \neq j$, q $\neq i$ and $v_q \in V_S$.

$R_j \geq 1$ means that the Market Capitalization of company j has not decreased. Therefore, the vertex v_j will keep the same state as before in the shareholding networks.

By the way, when $1 > R_j > x_3$, the assets of loss of holder q (denoted by m_q in (9)) can be obtained from the following equation:

$$m_q = \left(M_j - M'_j\right) \times u_{qj} \tag{10}$$

Using the network in Figure 1 as an example, if Holder 4 sells the stocks issued by company 2, the state of vertex 4 will be infected state. The corresponding influence on the networks can be depicted in Figure 5.

As shown in Figure 5(b), it will not impact on Company 2 when the proportion u_{ij} is less than 0.05 or Market Capitalization of company 2 has not decreased. However, if the Market Capitalization of company 2 shows a large decrease and $u_{42} \geq 0.05$, company 2 will go bankrupt (see Figure 5(c)). In addition, there is a middle state between Figures 5(b) and 5(c). Figure 5(d) shows that holder 4 selling the stocks issued by company 2 will make the Market Capitalization of company 2 decrease; i.e., the state of vertex 2 becomes infected state. It is worth noticing that company 2 can make a different decision according to its decrease values, marked by $C_3(v_2)$. Furthermore, as one of the holders of company 2, listed company 3 also will make a decision on account of the assets decrease, denoted by $H_2(v_3, m_3)$.

3.3.5. Decision Behavior of Listed Company j When Its Market Capitalization Decreases. Because the stock prices are the barometer of the economic development, which is sensitive to the general operating conditions, capital supply and demand, market demand, etc., generally speaking, the stock prices rise with the improvement of business performance. Therefore, the highly descent speed of Market Capitalization of listed companies means bad operation conditions. The phenomena of reduced cash flow or the fracture of capital chain may appear. At this time, the listed company may sell the holding stocks issued by some other listed companies to keep the company's normal operating cycle.

As mentioned in (9), if $0 \leq R_j \leq x_3$, the company j will go bankrupt or be delisted from the stock markets. So company j must take measures to avoid R_j approaching to x_3. Therefore, we can set a critical value x_4. If $x_3 < R_j \leq x_4$, the company j will make the corresponding adjustment to avoid going bankrupt or being delisted from the stock markets.

To sum up, the decision behavior of listed company j can be written as (11) when its Market Capitalization decreases.

$$C_3\left(v_j\right)$$
$$= \begin{cases} 0 & u_{jk} \equiv 0 \text{ or } R_j > x_4 \\ G - e_{jk} \text{ and } C_2\left(v_j, v_k, u_{jk}\right) & \exists u_{jk} > 0 \text{ and } x_3 < R_j \leq x_4 \end{cases} \tag{11}$$

where $k = 1, 2, 3, \cdots, n$ and $v_j, v_k \in V_L$.

The meanings of (11) are as follows.

(a) Initial state

(b) $u_{42} < 0.05$ or $R_2 \geq 1$

(c) $u_{42} \geq 0.05$ and $0 \leq R_2 \leq x_3$

(d) $u_{42} \geq 0.05$ and $x_3 < R_2 < 1$
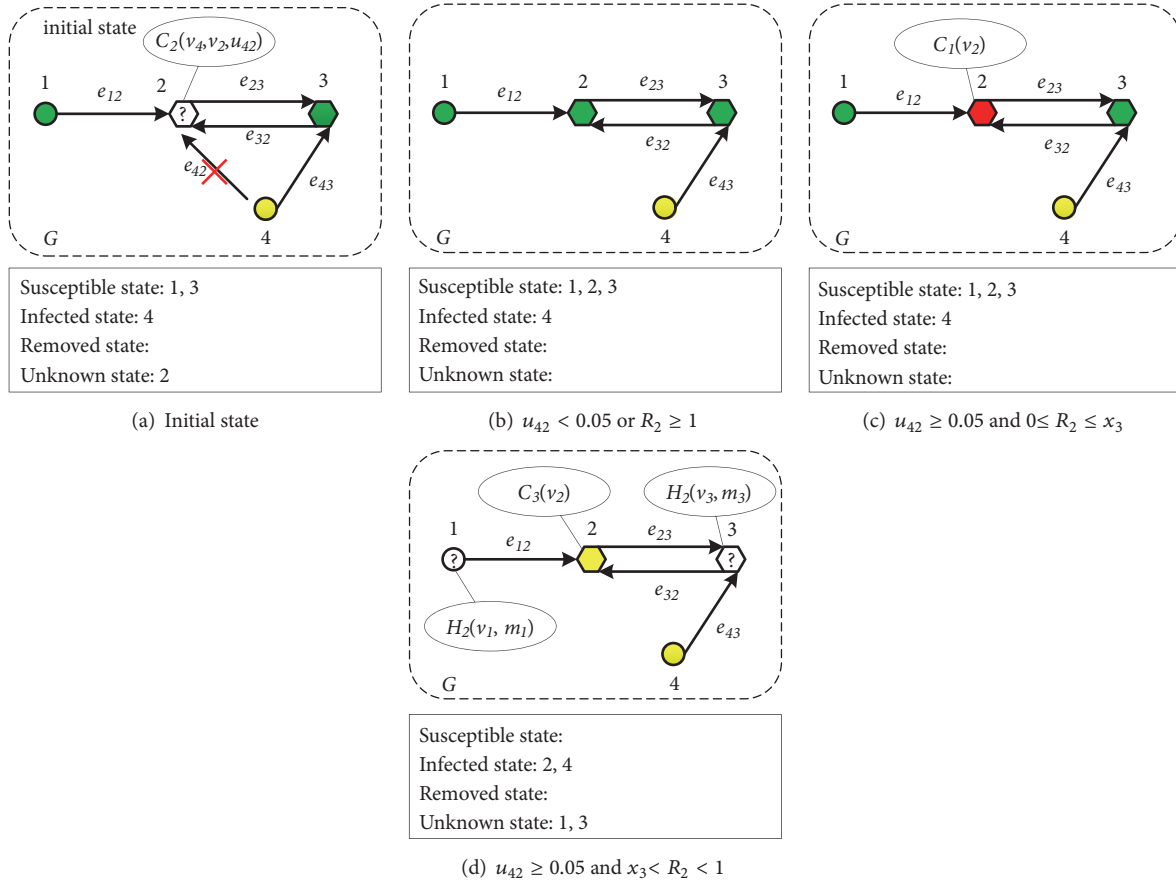
FIGURE 5: Influence of holder 4 on the networks when holder 4 sells the stocks issued by listed company 2.

If the listed company j does not hold the stocks issued by some other listed companies ($u_{jk} \equiv 0$) or the assets of loss are small ($R_j > x_4$), the company j will make no new decision and the vertex v_j keeps the same step as before in the shareholding networks. The vertex j will inherit the last state, such as vertex 2 in Figure 6(a).

If $x_3 < R_j \leq x_4$ and the listed company j holds stocks issued by some other listed companies, the listed company j will sell the stocks issued by company k according to the increase order of the assets which the company j invested to other companies. In the shareholding networks, the edge e_{jk} will be deleted and the influence of company j on listed company k when company j sells the stocks issued by listed company k, denoted by $C_2(v_j, v_k, u_{jk})$. For example in Figure 1, when the Market Capitalization of company 2 decreases, it will sell the stocks issued by company 3 in order to avoid critical funding shortages, as shown in Figure 6(b).

4. Characteristics of the Networks and Parameters Selection for the SIR Model

4.1. Characteristics of the Shareholding Networks. The Market Capitalization of Yearly End Date of the stock markets is ¥3.28×10^{13} in 2007 and ¥1.22×10^{13} in 2008. It denotes the Chinese Stock Markets have an obvious shrinkage

phenomenon because of the global financial crisis in 2008, which had a great influence on the Chinese economy. For the purpose of researching the robustness and vulnerability of Chinese stock markets against extreme circumstances, we will choose the data before and after the economic crisis to analyze. The data in 2007 is used as the sample in normal period; the data in 2009 is used as the sample of extreme circumstances after the crisis. As an example to simulate the crisis spreading in the networks, the data on December 31, 2007, is selected to establish the correlated networks between the listed companies and the main holders. The networks have 1534 listed companies, 13596 vertices, and 19326 edges.

First of all, we should judge the type of the networks. Thus, we did the linear regression analysis between the in-degree assets (or out-degree assets) and the cumulative probability value of the in-degree assets (or out-degree assets) of the vertices of the networks under double logarithmic coordinates. If we use γ as the symbol for the slope coefficients obtained by the linear regression analysis of the cumulative distribution, the probability distribution will follow a power law $p(k) \sim k^{-(-\gamma+1)}$ with the exponent $-\gamma + 1$. Thus, we use $-\gamma + 1$ as the symbol for the scale-free index [13, 16, 17]. Table 2 is the coefficients of linear fit in log–log scale.

The correlated coefficients approach to -1 as shown in Table 2, which indicates that all data points lie on a line

(a) $R_2 > x_4$

(b) $x_3 < R_2 \leqslant x_4$
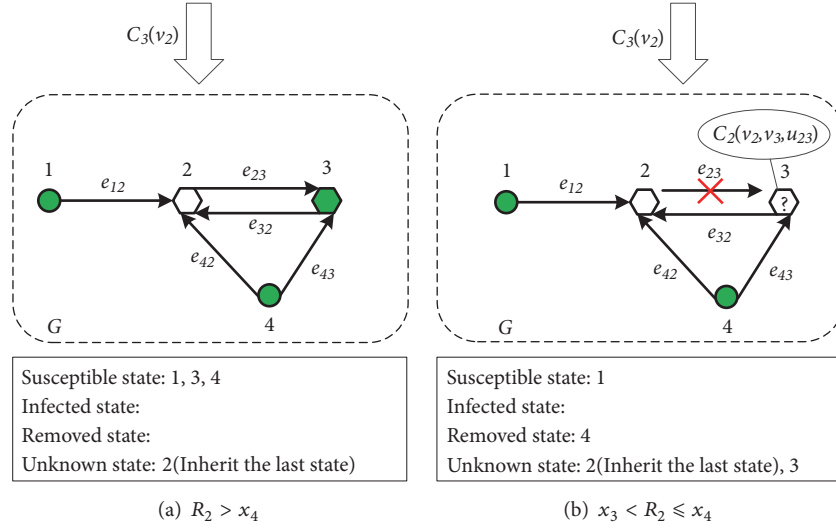
FIGURE 6: Decision behavior of listed company 2 when its Market Capitalization decreases.

TABLE 2: Correlation coefficients and the coefficients of linear fit in log–log scale.

Year	In-degree assets			Out-degree assets		
	Correlation	γ^{in}	Constant	Correlation	γ^{out}	Constant
2007	-0.9272	-1.3195	8.7319	-0.9655	-0.4200	2.7087
2008	-0.9283	-1.3255	8.6392	-0.9637	-0.4198	2.6787
2009	-0.9361	-1.2552	8.7380	-0.9580	-0.4214	2.7476

for which the cumulative probability value of the in-degree assets (or out-degree assets) decreases as in-degree assets (or out-degree assets) increases in log-log scale. Therefore, the networks belong to the scale-free networks [16–18].

The in-degree assets of the vertices mainly reveal the ability of the listed companies and funds attracting investors. As shown in Table 2, the indices γ^{in} range from -1.2 to -1.4; that is to say, the scale-free indices of in-degree assets of the vertices of the networks are between 2.2 and 2.4. It means that the in-hub vertices can only possess a small part of proportion [16]; i.e., the networks have little super listed companies. The in-hubs mainly represent some large listed companies in Sector I (Finance, Insurance).

The out-degree assets mainly represent the investments from main holders to listed companies and funds. According to Table 2, the indices γ^{out} range from -0.43 to -0.41, so the scale-free indices of out-degree assets of the vertices of the networks are between 1.41 and 1.43. Therefore, the distribution of the out-degree assets of the vertices of the networks are also similar to the distribution of the degree of the vertices in sparse scale-free networks. As we know, the mean value and variance of the cumulative distribution function of the degree in sparse scale-free networks are divergent. It means that there are relatively more out-hub vertices in the networks, which mainly represent two kinds of companies: (1) some companies in sector I (Finance, Insurance), such as the state-owned commercial banks and insurance companies and (2) some parent company owning several listed companies. For example, the Aviation Industry

Corporation of China has more than 20 subsidiary companies listing on the stock markets.

4.2. Parameters Selection for the SIR Model through Empirical Research. In the RESSET Financial Research Database, the number of stocks with complete information of Market Capitalization of Yearly End Date is 1636 in 2007 and 1804 in 2009. The number of stocks with complete information of Market Capitalization of Yearly End Date both in 2007 and in 2009 is 1628. The Market Capitalization of Yearly End Date of a certain listed company j in 2009 divided by its Market Capitalization of Yearly End Date in 2007 is denoted by δ. Then we get the scatter diagram of cumulative distribution of δ, as shown in Figure 7.

According to Figure 7, the index δ ranges from 0.239 to 79.598. Most of the points range from 0.4 to 2.4. The point with maximum value of 79.598 represents Qinghai Salt Lake Industry Group Company Limited (stock code: 000578). By the way, δ is equal to 79.589, which is not a normal phenomenon, and the well-known case about salt lake 4.4 billion Yuan equity is always in inquisition stage until now.

On May 8, 2003, Shanghai Stock Exchange and Shenzhen Stock Exchange started a warning mechanism for stocks that is incurring risk of being removed. This is an extra treatment of the Special Treatment Mechanism, and the original stock name will be prefixed with '∗ST'. The daily price up and down limit is also 5%. Furthermore, if the '∗ST' stocks continue to make a loss the next year, they will be temporarily delisted.
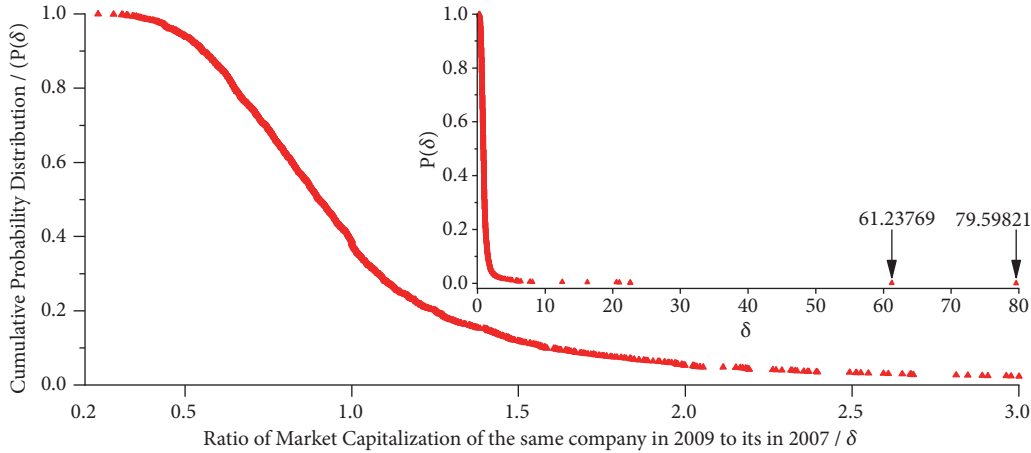
FIGURE 7: Cumulative probability distribution of δ.

TABLE 3: Fitting parameters of Market Capitalization of Yearly End Date.

Optimization Algorithm	$0.05 \leq f < 0.5$		$0.5 \leq f$	
	Differential Evolution	Levenberg-Marquardt	Differential Evolution	Levenberg-Marquardt
Mean square error	0.204586	0.204586	0.150582	0.150582
Root mean squared error	11.133538	11.133545	0.453501	0.453500
R square	1.000000	1.000000	1.000000	1.000000
a	0.959815	0.959815	0.964046	0.964079
b	-0.013473	-0.013474	-0.054734	-0.055387

Checking the original data, we find that 24 listed companies are prefixed with '∗ST'. 8 listed companies are delisted from the stock markets (stock code: 000515, 000569, 420058, 600001, 600357, 600627, 600786, and 600840). The data of Market Capitalization of Yearly End Date in 2009 of them are not found in the RESSET database. We suppose that the delisted companies are the worst operating companies with highly descent speed of Market Capitalization of Yearly End Date. Therefore, x_3 in (9) is equal to 0.239. Suppose the operating status of the 34 listed companies prefixed with '∗ST' is worse than the other listed companies in the stock markets; we can get the critical value x_4 in (11), which will be equal to 0.419.

If a holder only holds the stocks issued by one listed company, the holding market value of the holder will become 0.419 times of that in the previous. Under this condition, the best way to decrease the holder's losses is selling the stocks. Therefore, the critical value x_2 in (7) will be equal to 0.419.

For x_1 in (7), Martin Zweig has suggested holders should sell the stocks when its price declines by 10%~20% of Bid Price. So we set $x_1 = 0.2$ in this paper [19].

As mentioned above, we have proved that the probability of the weighted degree of vertices, which represents the Market Capitalization of stocks, followed the power-law. Therefore, we suppose newly Market Capitalization M'_j, previous Market Capitalization M, and the proportion of single transaction shares of a certain listed company (denoted by f) have the following relationships.

$$\lg M' = a \times \lg M + b \times f \qquad (12)$$

where a, b are the parameter.

As mentioned above, the data before and after the Subprime Crisis are selected to analyze the crisis spreading in the networks. Now, suppose the Market Capitalization of Yearly End Date in 2007 is M and in 2009 is M'. The block trade of stocks with proportion greater than 5% in RESSET database is used as the sample data. The number of valid sample data sets is 286. The number of sample data sets with $0.05 \leq f < 0.5$ is 266, and the number of sample data sets with $0.5 \leq f$ is 20. Two methods have been used to find out a, b. The results are listed in Table 3.

As shown in Table 3, the R square approaches to 1 even retaining 6 decimal places. It denotes that the block trades of stocks with proportion greater than 5% can induce the listed companies' Market Capitalization fluctuation. Meanwhile, (12) just can reveal the changes.

According to the analysis shown in Table 3, we set the a = 0.959815, b = −0.013473 when $0.05 \leq f < 0.5$, and a = 0.964046, b = −0.054734, when $0.5 \leq f$.

5. Numerical Simulation of Crisis Spreading

The MATLAB program is used to simulate the crisis spreading in the shareholding networks under the condition of the networks meeting random failure or intentional attack. The random failure refers to removing the vertices randomly. The intentional attack refers that the vertices are removed from big to small according to the degree of the vertices.

The number of initial failure vertices ranges from 0 to 500. When the number of the initial failure vertices is equal to 500, it includes 367 listed companies and 133 nonlisted holders under the condition of the networks meeting intentional
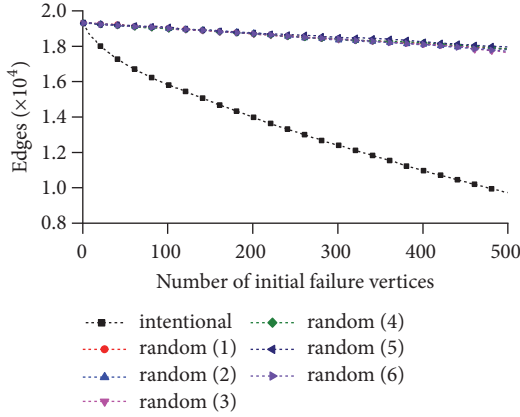
FIGURE 8: Relationships between the edges and the initial failure vertices in the networks.
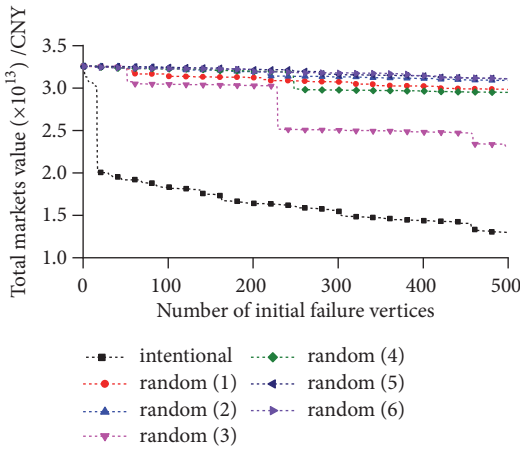
FIGURE 9: Relationships between the total markets value and the initial failure vertices in the networks.

attack. As for the analysis of crisis spreading under the condition of the networks meeting random failure, we select the six sample data, including 48, 56, 57, 51, 52, or 58 listed companies, respectively.

After carefully analyzing, the relationships between the edges and the initial failure vertices in the networks are plotted in Figure 8. The relationships between the total Market Capitalization and the initial failure vertices in the networks are shown in Figure 9.

The random failure can reveal that a handful of listed companies go bankrupt or nonlisted holders are delisted from the stock market under normal circumstances. As shown in Figures 8 and 9, we note that the crisis does not widely spread in the stock markets under this condition. The networks as a whole have good robustness to the random failures. Even with 500 initial failure vertices, the total Market Capitalization of stock markets has declined a little.

The intentional attack can reveal the breakdown of a handful of hub vertices which represents large-scale listed companies and holding companies going bankrupt. Under this condition, the crisis has widely spread in the stock markets and produced quite a lot of damage. There is highly

descent speed of total Market Capitalization and the edges of the networks (see Figures 8 and 9). The network has changed obviously and collapsed at a tremendous speed when the number of the initial failure vertices reaches 15. It denotes that the networks have obvious vulnerability to the intentional attacks. The breakdown of Fannie Mae and Freddie Mac which caused the financial crisis in 2007 is similar to intentional attack in the shareholding networks.

When networks meet random failure and intentional attack, the number of the susceptible state, infected state, and removed state of nonlisted shareholders is listed in Figure 10. The number of the susceptible state, infected state, and removed state of listed companies is listed in Figure 11.

Generally, both Figures 10(a) and 11(a) show that the number of susceptible vertices representing nonlisted holders and listed companies under the condition of the networks meeting random failure is more than that under the condition of the networks meeting intentional attack, respectively. Both Figures 10(c) and 11(c) show that the number of removed vertices representing nonlisted holders and listed companies under the condition of the networks meeting random failure is less than that under the condition of the networks meeting intentional attack, respectively. These phenomena coincide with the phenomena revealed in Figures 8 and 9; i.e., the networks have good robustness to the random failures and the obviously vulnerability to the intentional attacks. Such characters accompany with the existence of the highly linked hub vertices in scale-free networks, because intentional attack means removing the vertices beginning with the biggest hub vertices and the robustness of the networks can be destroyed easily under the intentional attacks. As mentioned in Ma e Zhuang *et al.*, 2011, and Section 4, the hub vertices mainly represent the super-scale state-owned enterprises, commercial banks, and insurance companies. The closed mutual investment relationships among the hub vertices make them group rich-club spontaneously. There is little linkage between the members of rub-club and the other vertices. Thus, the network shows obviously catastrophic phenomenon. It can just explain why Figures 8 and 9 present a breakpoint.

As proved in Section 4, the networks belong to the scale-free networks in which the linkages among vertices are different. Most of the listed companies have not closely interacted together directly. Thus, the number of listed companies at infected state is small (see Figures 10(b) and 11(b)) either the network meeting random failure or intentional attack. As for the intentional attack on the network, it means that the members of hub-club have been attacked. Under this condition, the network will take on the phenomenon that the same listed companies have been attacked repeatedly because of the closely linkages among the members of hub-club. Therefore, the number of the vertices at infected state is relatively small.

The coordinate axes of Figures 10 and 11 are uniformed by a unified way. Moreover, it is worthy to note that the Y-axis scales (multiplied by 10^2 and 10^3, respectively) are ten times different, respectively. Actually, more than 88% vertices in the shareholding networks represent nonlisted holders. Therefore, Figure 11 can reveal the overall changing trends of
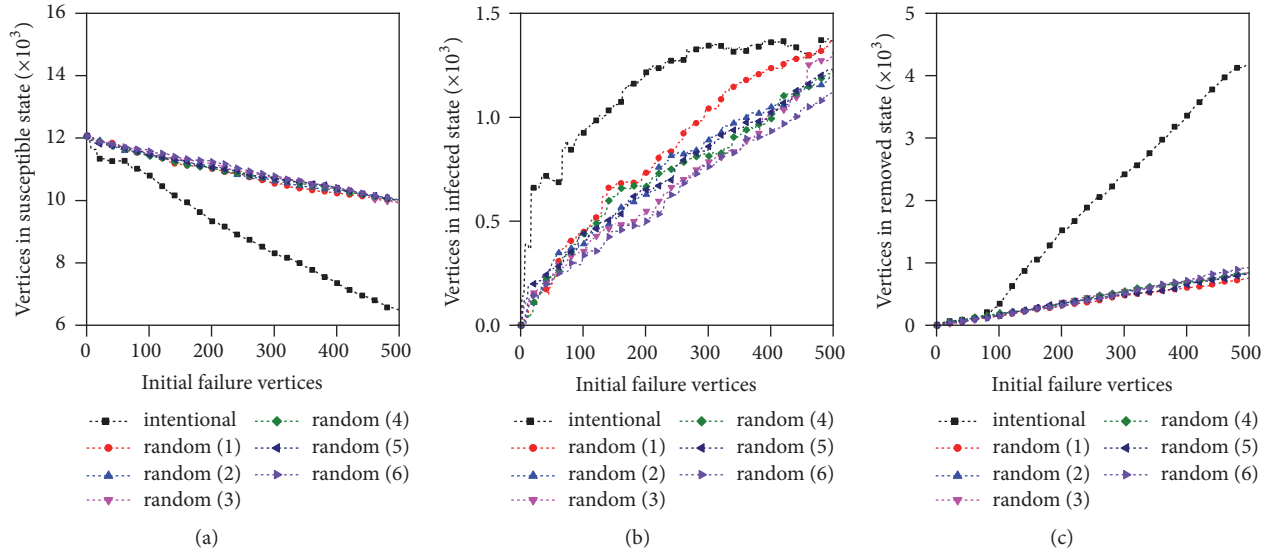
FIGURE 10: Crisis spreading when networks meet random failure and intentional attack (nonlisted shareholders).
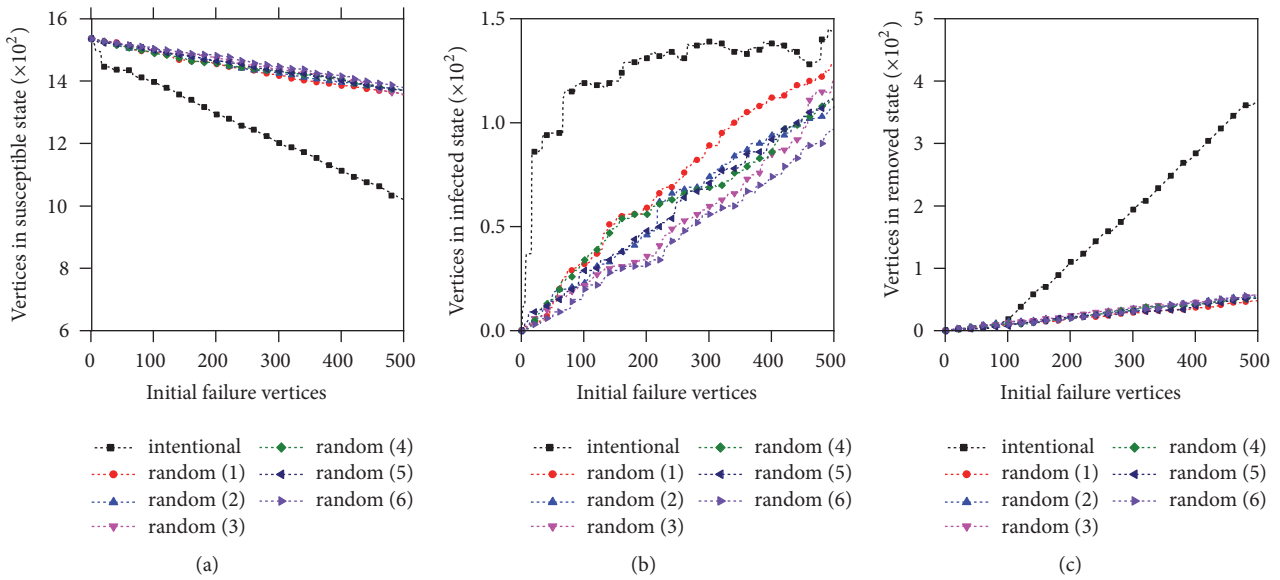
FIGURE 11: Crisis spreading when networks meet random failure and intentional attack (listed companies).

susceptible vertices, infected vertices, and removed vertices of the networks.

When the number of initial failure vertices is less than 5, the number of susceptible (infected or removed) vertices representing nonlisted holders and listed companies under the condition of intentional attack is similar to that under the condition of the networks meeting random failure, respectively (see Figures 10 and 11); i.e., a small quantity of hub vertices removing from the networks has not decreased the scale of the networks obviously. Moreover, the crisis has not widely spread in the networks when the number of initial vertices is small. It is mainly because the large-scale listed companies' holders are also handing some

other listed companies' stocks. They have good tolerance to the failures in stock markets when we only consider the capital chain between the listed companies and main holders and neglect the rumor spreading in complex networks.

In addition, there is an interesting phenomenon emergence that the fat-tailed degree distribution diverges as the growth number of removed vertices when the networks meet intentional attacks. The scale-free properties are not preserved due to removing the hub vertices continually. Just as Moore C. et al. [20] and Piccardi C. et al. [21] point that the scale-free properties may not preserve in the long run if there is disappearance or death of vertices.

TABLE 4: Controllability of the shareholding networks under attacks or failures.

Number of initial failure vertices	Intentional Attack				Random Failure			
	n_D	n_{Ds}	n_{Dc}	n_{DC}	n_D	n_{Ds}	n_{Dc}	n_{DC}
0	0.8897	0.9969	0.0031	0.0241	0.8897	0.9969	0.0031	0.0241
50	0.8892	0.9969	0.0031	0.0242	0.8894	0.9969	0.0031	0.0240
100	0.8879	0.9970	0.0030	0.0231	0.8891	0.9969	0.0031	0.0245
150	0.8854	0.9969	0.0031	0.0232	0.8887	0.9969	0.0031	0.0243
200	0.8837	0.9967	0.0033	0.0245	0.8884	0.9968	0.0032	0.0245
250	0.8824	0.9964	0.0036	0.0260	0.8880	0.9968	0.0032	0.0245
300	0.8808	0.9966	0.0034	0.0246	0.8876	0.9968	0.0032	0.0245
350	0.8795	0.9964	0.0036	0.0254	0.8872	0.9968	0.0032	0.0248
400	0.8774	0.9963	0.0037	0.0256	0.8869	0.9968	0.0032	0.0247
450	0.8753	0.9962	0.0038	0.0258	0.8866	0.9968	0.0032	0.0244
500	0.8741	0.9961	0.0039	0.02668	0.8862	0.9967	0.0033	0.0247

6. Controllability of Crisis Spreading

As mentioned above, random failures and intentional attacks have striking difference. It has close similarities with dual strategies of targeted vaccinations for controlling the spread of infectious diseases [22, 23]. For the purpose of discussing the difference of controllability of shareholding networks, we have studied the controllability and controlling strategies of the shareholding networks when the networks meet random failures or intentional attacks on basis of structure controllability theory [8], which is extracted from mapping maximum matching for networks and Kalman's controllability rank condition. The controllability of the networks under the condition of intentional attacks and random failures is listed in Table 4.

In Table 4, n_D denotes the controllablility of the networks, which is the ratio of the number of minimum number of driver vertices (denoted by N_D) to the total number of the vertices in the network. n_{Ds} and n_{Dc} are ratio of the number of driver vertices representing nonlisted holders or listed companies to N_D, respectively. n_{DC} is the number of driver vertiecs representing the listed companies to the total number of verties representing the listed companies.

As shown in Table 4, the proportion of minimum number of driver vertices n_D is as high as 87%, including the vertices representing nonlisted holders about 99%. Such a result seems unreasonable, but we think and consider how things happen in this way. We should note that this theory of controllability pays more attention on controllability the whole networks. Of course, to achieve such an effect, the more driver vertices involve in it, the better results networks get. More than 99% driver vertices belong to nonlisted holders which means that controlling the whole stock markets relies on some input signals which can act on whole stock markets directly or can induce holders decision-making in an indirect way. Maybe a politically good new or an appropriate exciting economical policy can play the key role together with a proper guide of the public opinion when the markets meet failures or attacks.

In addition, it is worthy to note that the proportion of driver vertices representing listed companies is less than

0.4%, and it is also less than 3% of the total number of the listed companies in the networks. So, a question worth thinking deeply is how well it works when only controlling the driver vertices representing listed companies.

To illustrate this problem, the concept of global efficiency in complex networks is introduced here. It is an associated concept with the average path length of the network in graph theory. When the distance between two vertices is shorter, the efficiency of transmitting information between them is higher; that is, the efficiency of transmitting information between two vertices is proportional to the reciprocal of the distance between them. The average efficiency of all vertices in the network can reflect the average efficiency of information sent between vertices in the network. The efficiency of the network (indicated by η_G) and the average efficiency (indicated by η_C) of the listed company in the driver vertices are given in Table 5.

However, under normal circumstances, controlling a small part of driver vertices representing listed companies, applying appropriate control strategies, and using its characteristics of high efficiency of sending information, can effectively control the stock market. Our research provides a new reference to further exploration about the transmission mechanism of the crisis based SIR model and further research on the controllability of crisis spreading in financial markets.

As can be seen from Table 5, the global efficiency of the network is significantly lower. However, it is worth noting that η_C is not only obviously high, but also gradually reduced under intentional attacks. This phenomenon is not obvious under random failures. In order to clearly illustrate this phenomenon, the ratio of η_C to η_G is given in Table 5. This means that intentional attacks obviously cause substantial damage to the hub vertices in the network with listed companies as the core. However, the damage caused by random attacks is not obvious, and in the case of random attacks, the ratio between η_C and η_G is more than 1000 times. Combined with Table 4, this clearly reveals that the control of 2%-3% driver vertices representing listed companies, the application of appropriate control strategies, and the use of its high efficiency of sending information can effectively control the trends of stock market to a certain degree in case of random failures.

TABLE 5: Efficiency of the shareholding networks under attacks or failures.

Number of initial failure vertices	Intentional Attack			Random Failure		
	η_G	η_C	η_C/η_G	η_G	η_C	η_C/η_G
0	2.411E-04	0.2908	1206.1	2.411E-04	0.2908	1206.1
50	1.982E-04	0.1554	784.1	2.420E-04	0.2877	1189.0
100	1.902E-04	0.1427	750.0	2.432E-04	0.2864	1177.6
150	1.993E-04	0.1315	659.8	2.443E-04	0.2849	1166.0
200	2.044E-04	0.1249	611.3	2.459E-04	0.2827	1149.8
250	2.096E-04	0.1216	580.0	2.464E-04	0.2764	1121.6
300	2.134E-04	0.1007	471.9	2.476E-04	0.2720	1098.5
350	2.198E-04	0.0968	440.5	2.492E-04	0.2716	1089.5
400	2.293E-04	0.0932	406.4	2.502E-04	0.2693	1076.2
450	2.420E-04	0.0911	376.4	2.507E-04	0.2652	1057.5
500	2.465E-04	0.0794	322.1	2.514E-04	0.2622	1042.7

Just as shown in Table 5, the small part of driver vertices representing listed companies, which mainly denote some large-scale listed companies, has high efficiency of sending information ability. Thus, the stock prices of many other listed companies are highly influenced by the stock price of large-scale listed companies. Meanwhile, as mentioned in [13], the highly linked hub vertices mainly represent the super-scale state-owned enterprises, commercial banks, and insurance companies. These companies are indispensable for our life. In addition, in our networks, the nonrational decisions and some other existing nonlinear factors are neglected. Thus, once the large-scale companies go bankrupt, the rumor among the ordinary shareholders and some other unknown factors will accelerate the spreading of the crisis. The bubble in the stock markets will collapse easily. Then the phenomenon of domino effect may appear. Correspondingly, a financial crisis may take place that the US subprime mortgage crisis in 2007 triggering the worldwide financial crisis is just a good case in point. Therefore, when some large-scale listed companies go bankrupt, the government must do its best to avoid the rumor spreading in the stock markets and prevent the catastrophe.

7. Conclusion

In this paper, the mutual influences between listed companies and their main holders because of the broken financing chain are studied to reveal the crisis spreading in shareholding networks of listed companies and their main holders. The crisis-spreading model of susceptible-infected-removed (SIR) is established. The numerical computation method has been successfully used to analyze the crisis spreading in the shareholding networks and its controllability when the networks meet random failures or intentional attacks. The main conclusions are follows.

The crisis spreads at a rapid speed and the total Market Capitalization has obviously decreased when the networks meet intentional attack. It means that the intentional attack on hub vertices produces quite a lot of damage. The overall trends of crisis spreading in the networks can be viewed through the changes of the number of vertices in susceptible

state, infected state, and removed state. The descent speed of the number of susceptible vertices when networks meet intentional attack is faster than that when networks meet random failure. The rising speed of the number of infected vertices (and removed vertices) when networks meet intentional attack is faster than that when networks meet random failure. Thus, the networks meeting intentional attack show more obviously vulnerability than the networks meeting random attack. It mainly attributes to the enterprises in the sector of Finance and Insurance, and some super-scale companies. The relationships among these enterprises and the other large-scale listed companies or the holding companies are closed. Therefore, intentional attack has huge damage on the shareholding networks. At the same time, these companies are indispensable for our life. Once these companies go bankrupt, the rumor will accelerate the crisis spreading. Then the phenomenon of domino effect may appear easily. Correspondingly, an economic crisis takes place.

Finally, The network-based structural controllability theory conducts controllability research on the network when it is subjected to intentional attacks and random failures. The research shows that if the network needs to be controlled globally, the policy-oriented role and the public opinion control strategy should be used in the crisis. Under normal circumstances, controlling a small part of driver vertices representing listed companies, applying appropriate control strategies, and using its characteristics of high efficiency of sending information can effectively control the stock market. In this way, the government can avoid to get into the financial whirlpool and speed a lot of funds on relieving the initial failure listed companies.

8. Further Discussion

As mentioned above, more than 70% of the Market Capitalization of Chinese stock markets has vaporized during the year of 2008. At the same period, the maximum falling range of Hang Seng Index is 65% during the year of 2007 and 2008, where the total Market Capitalization of A-shares issued by some state-owned listed companies including 600019, 600028, 601088, 601857, 601628, 601318, 601600, 601988,

601601, 601919, and 601111 has vaporized ten trillion CNY since the end of 2007. Why did so serious vaporized phenomenon of state-owned enterprises appear? Now, we need to recall that the purpose of founding the Chinese mainland stock markets is to solve the financing difficulties of state-owned enterprises. The Market Capitalization of Chinese mainland stock markets is 26 trillion CNY on December 2011, while the Market Capitalization of state-owned enterprises is 20.3 trillion CNY. Thus, the fluctuation trends of Chinese mainland stock markets are under the influence of state-owned enterprises deeply.

Hong Kong stock markets are developed stock markets and they have a relatively better risk resistance capacity than mainland stock markets because of highly opening level and reasonable economic structure. Just as Justin Yifu Lin (former Chief Economist and Senior Vice President of the World Bank.) points that many problems in Chinese financial system are caused by the problem of state-owned enterprises, we also believed that the poor risk resistance capacity of Chinese financial system and the current low levels of Chinese mainland stock markets mainly attribute to the numerous large-scale listed companies and large-scale holding companies in Chinese stock markets.

Perhaps only when we solve the problems left over by history about state-owned enterprises, can we improve the risk resistance capacity of Chinese financial system. The initial thought of this paper is to establish a crisis spreading model to find out effective strategies for controlling the widely spreading crises in stock markets and to come up with some policy suggestions for the healthy development of Chinese stock markets. It is a pity that the current model established in this paper cannot simulate the crisis spreading in Chinese stock markets accurately. Due to some variables of the model needing further confirmation and many realistic circumstances needing consideration, such as the investors' sentiment and investment behavior, the public opinions, and some other macroeconomic factors, this model is only a simplified academic model.

However, it is worth celebrating that this method discussed in this paper gives us a possible way yet to explore the crisis speeding mechanism and its controllability. In addition, the SIR model proposed in this paper is also beneficial to master the transmission mechanism of crisis spreading on the stock markets and decrease the loss of the economic entity. Further research on the SIR model can help to reveal the dissemination process and root cause of financial crisis.

Data Availability

The data in text format or EXCEL format used in our manuscript to support the findings of this study are extracted from the RESSET Financial Research Database (http://www.resset.cn/), including all the stocks issued in the Shanghai Stock Exchange and the Shenzhen Stock Exchange before December 31, 2009. The documents we selected include the Main Stockholders List and Ownership Structure and the Yearly Market Capitalization (CNY) of all listed companies. The database was supplied by Beijing Gildata Resset Data Tech Co., Ltd., under license and so cannot be made freely available. Requests for access to these data should be made to Beijing Gildata Resset Data Tech Co., Ltd. (address: B502-503, Caizhi Building, Zhongguancun East Road No. 18, Haidian District, Beijing, China. Tel: +86-10-82601461).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.

[2] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.

[3] A. Ramani, A. S. Carstea, R. Willox, and B. Grammaticos, "Oscillating epidemics: a discrete-time model," *Physica A: Statistical Mechanics and its Applications*, vol. 333, no. 1-4, pp. 278–292, 2004.

[4] C. Piccardi and R. Casagrandi, "Inefficient epidemic spreading in scale-free networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, no. 2, 2008.

[5] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, "Epidemic outbreaks in complex heterogeneous networks," *The European Physical Journal B*, vol. 26, no. 4, pp. 521–529, 2002.

[6] Y. Moreno, J. B. Gómez, and A. F. Pacheco, "Epidemic incidence in correlated complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 3, 2003.

[7] Y. Moreno, M. Nekovee, and A. F. Pacheco, "Dynamics of rumor spreading in complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 6, Article ID 066130, 2004.

[8] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–173, 2011.

[9] T. Jia and A. Barabási, "Control Capacity and A Random Sampling Method in Exploring Controllability of Complex Networks," *Scientific Reports*, vol. 3, no. 1, 2013.

[10] T. Jia, Y. Liu, E. Csóka, M. Pósfai, J. Slotine, and A. Barabási, "Emergence of bimodality in controlling complex networks," *Nature Communications*, vol. 4, no. 1, 2013.

[11] Z. Z. Yuan, C. Zhao, Z. R. Di, W.-X. Wang, and Y.-C. Lai, "Exact controllability of complex networks," *Nature Communications*, vol. 4, 2013.

[12] J. Sun and A. E. Motter, "Controllability Transition and Nonlocality in Network Control," *Physical Review Letters*, vol. 110, no. 20, 2013.

[13] Y.-Y. Ma, X.-T. Zhuang, and L.-X. Li, "Research on the relationships of the domestic mutual investment of China based on the cross-shareholding networks of the listed companies," *Physica*

A: Statistical Mechanics and its Applications, vol. 390, no. 4, pp. 749–759, 2011.

[14] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, Macmillan Press, New York, NY, USA, 1976.

[15] "Rules Governing the Listing of Stocks on Shanghai StockExchange," http://english.sse.com.cn/laws/framework/c/4547752 .pdf. (Revised in 2018).

[16] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.

[17] Y. Ma and L. Li, "Research on the topology, community and robustness of the correlated networks of listed companies, funds and their main holders in China," in *Proceedings of the 2012 International Conference on Information Management, Innovation Management and Industrial Engineering, ICIII 2012*, pp. 40–46, China, October 2012.

[18] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.

[19] M. Zweig, *Martin Zweig's Winning on Wall Street*, Grand Central Publishing, 1997.

[20] C. Moore, G. Ghoshal, and M. E. J. Newman, "Exact solutions for models of evolving networks with addition and deletion of nodes," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 3, Article ID 036121, 2006.

[21] C. Piccardi, A. Colombo, and R. Casagrandi, "Connectivity interplays with age in shaping contagion over networks with vital dynamics," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 91, no. 2, 2015.

[22] R. Cohen, S. Havlin, and D. Ben-Avraham, "Efficient immunization strategies for computer networks and populations," *Physical Review Letters*, vol. 91, no. 24, Article ID 247901, 2003.

[23] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge, UK, 2008.

Research Article

Complexity of a Microblogging Social Network in the Framework of Modern Nonlinear Science

Andrey Dmitriev [iD], Vasily Kornilov, and Svetlana Maltseva

School of Business Informatics, National Research University Higher School of Economics, Moscow 101000, Russia

Correspondence should be addressed to Andrey Dmitriev; a.dmitriev@hse.ru

Recent developments in nonlinear science have caused the formation of a new paradigm called the paradigm of complexity. The self-organized criticality theory constitutes the foundation of this paradigm. To estimate the complexity of a microblogging social network, we used one of the conceptual schemes of the paradigm, namely, the system of key signs of complexity of the external manifestations of the system irrespective of its internal structure. Our research revealed all the key signs of complexity of the time series of a number of microposts. We offer a new model of a microblogging social network as a nonlinear random dynamical system with additive noise in three-dimensional phase space. Implementations of this model in the adiabatic approximation possess all the key signs of complexity, making the model a reasonable evolutionary model for a microblogging social network. The use of adiabatic approximation allows us to model a microblogging social network as a nonlinear random dynamical system with multiplicative noise with the power-law in one-dimensional phase space.

1. Introduction

Social networks have been studied longer than any other type of networks. It is remarkable that one of the signs of network complexity—a power law of nodes' degree distribution [1]—was first empirically formulated by D. Price in 1965 for social networks. In 1999, A. L. Barabasi, a physicist from the University of Notre Dame (USA), and his graduate student R. Albert determined [2, 3] that, for many networks, instead of the expected Poisson probability distribution of nodes' degree (i.e., the number of connections a node has to other nodes), the distribution they obtained approximately followed a power law as all critical states do. In many real networks, a small number of nodes have a large number of connections, whereas a large number of nodes have just a few connections. Such networks are called scale-free networks. This name was not invented specifically for this type of networks. It came from the theory of critical phenomena, where fluctuations in critical states also follow a power law. The theory of scale-free networks is considered to be one of the scenarios complex systems follow when they come into a critical state. As of late, such networks are more often called complex networks.

Some other relevant works in this area are those of refs. [4–8].

An extensive body of research on the modeling of the structure and functioning of social networks is available today. This research has two directions. The first direction relates to the analysis of the social networks data (see one of the latest reviews [9]), while the second concerns the development of models of the structure, dynamics, and evolution of social networks. The distinction between these two directions is somewhat arbitrary, since in most cases these directions overlap (see, e.g., [10, 11]).

Starting from the second half of the 20th century, the ideas and methods of physics have tended to infiltrate natural sciences and traditional humanities. Methods of physical modeling are often used in such areas of science as demographics, sociology, and linguistics. As a result, sociophysical models of social networks, such as the Ising model [12–15], Bose-Einstein condensate model [3, 16], Quantum walk model [17], Ground state and community detection [18, 19], among others, were developed.

Despite having a variety of sociophysical models, the results and theories of nonlinear science, with some

exclusions (see, e.g., [20, 21]), are not used to model the evolution of social networks. First of all, we are talking about the complexity and self-organized criticality theory describing the mechanism of complexity [22–24]. Mechanisms of self-organized criticality in social knowledge creation process are presented in the paper [25]. It is noteworthy that the key sign of complexity of a system regardless of its internal structure, i.e., one based solely on its external characteristics, was formulated in the framework of this theory. According to this theory, a system is considered to be complex if it is able to generate unexpected and/or extraordinary events (for instance, bursts of values in time series). This motivated our research. The purpose of the research is a nonlinear dynamical interpretation of the complexity of a microblogging network and the development of an appropriate network model that could explain its complexity using the third paradigm of nonlinear science called the complexity paradigm. Another motivation for the research was the results presented in [26–31] where the time series of a number of microposts are characterized by the majority of key signs of the system complexity (a detailed description of the key signs of the system complexity is presented in Section 2).

This paper is organized as follows. Section 2 deals with the key signs of the system complexity according to the complexity paradigm. Section 3 presents the results of the analysis of an empiric time series of a number of microposts, including the results of the calculation of the key signs of the complexity. Section 4 presents a model of a microblogging social network as a nonlinear deterministic dynamical system including its capabilities and restrictions. Section 5 presents a generalized model of a microblogging social network, modified by the consideration of stochastic sources and a decrease in the order parameter, as well as the results of an analysis in the adiabatic approximation. Section 6 contains the main results of the research and a discussion.

2. Nonlinear Dynamical Interpretation of Complexity

The development of any branch of science leads to the formulation of paradigms, namely, initial conceptual schemes, models of problem statements, and solutions of the problems. At this time, three paradigms have been developed in nonlinear science. The first paradigm is that of self-organization. The second is the paradigm of deterministic chaos. The most recent development of nonlinear science is closely linked to the third paradigm, which could be defined as a paradigm of complexity that has the theory of self-organized criticality as its foundation. The paradigm of complexity lies at the junction of the first two paradigms. If the first two paradigms deal with order and chaos, respectively, the third is usually described as "life on the edge of chaos" [32].

Since it is impossible to rigorously define complexity, our research is limited to consideration of the key signs of system complexity defined in the publications by Per Bak and co-authors [22–24], and their application to the interpretation of the complexity of microblogging social networks. As stated in the introduction, first of all, we consider the complexity of external system manifestations regardless of internal structure. For the purposes of this research, we define "external system manifestations" as signals (the time series of a number of microposts) of a microblogging social network generated as a result of nontrivial interactions within a very large pool of users.

One of the key signs of complexity is its inclination to the occurrence of catastrophic events—either unexpected (i.e., nonpredictable) or extraordinary (i.e., prominent among similar events), or both. Importantly, in either case we can conclude that the system that has generated such an event is complex. From simple systems, we could expect predictability and similarities in their behavior. As for the signals of a microblogging system, such events qualitatively correspond to considerable bursts seen on a plot of value increments of the time series of a number of microposts. One of the quantitative criteria of the existence of catastrophic events is the existence of power low of the probability density function (PDF) for the values of the time series. It is worth mentioning that, in the majority of cases, the occurrence of such events on the network signal level corresponds to the qualitative restructuring of the system, i.e., a transition from a polycentric state to a monocentric state, and vice versa (such transitions are thoroughly described in [33]).

Another key sign of complexity is scale invariance, meaning that events or objects lack their own characteristic dimensions, durations, energies, etc. At the level of external manifestations of a microblogging network, scale invariance means that the time series of a number of microposts are fractal or multifractal time series (such time series are described in detail in [34]).

In a general case, a power low for PDF is a statistical expression of scale invariance of the time series:

$$p(x) \propto x^{-\alpha}, \tag{1}$$

where usually $\alpha \in (1, 2)$. PDF (1) belongs to the class of fat-tailed PDFs. For statistical description of catastrophic events, PDF (1) is a rule with almost no exceptions. PDF (1) differs from compact distributions (for example, Gaussian distribution) because the events corresponding to the tail of the distribution are not rare enough to be neglected. PDF (1) reflects a strong interdependence of the events. For example, such distribution may be caused by an avalanche-like increase of the number of microposts in the network as a result of a "chain reaction" caused by reposting.

Another manifestation of the scale invariance of the time series is the existence of the power spectral density (PSD) specific for flicker noise:

$$S(f) \propto f^{-\beta}, \tag{2}$$

where $\beta \cong 1$. The existence of PSD (2) means that a considerable part of the energy is linked to very slow processes. For a microblogging network, the existence of PSD (2) means that it is impossible to predict the behavior of the time series of a number of microposts without considering global information exchange processes.

The aforementioned features of PDF and PSD are not the only criteria of scale invariance. Besides PDF and PSD,

we used a fractal dimension and a Hurst exponent along with other quantitative measures and criteria. It is important to stress that the scale invariance and an inclination to catastrophes are typical only for systems that are far from equilibrium. Therefore, a nonequilibrium state of the system and, therefore, a nonlinearity are the necessary conditions for the complexity of the system.

Lastly, the third key sign that characterizes complex systems is their integrity. The integral properties of a system usually are statistically described by power-law space and time correlations. These correlations are known as distant space and time correlations. The existence of distant time correlations or long memory in time series is characterized by the autocorrelation function (ACF) in the following form [34]:

$$ACF_\tau \propto \tau^{-\gamma}, \tag{3}$$

where $\gamma \in (0.1)$. The existence of the relationship (3) implies the absence of characteristic times at which the information about the previous events could be lost. A catastrophic behavior and integrity are connected in the following way: for the catastrophic behavior, part of the system should be able to function in coordination. For a microblogging network, an avalanche-like increase of the number of microposts is possible when a user and his followers, followers of these followers, etc., are working in coordination. Integrity is possible in complex systems only due to the processes of self-organization. Here we talk about coarse scale properties of the system, since minor changes in system parameters do not affect its integrity.

Therefore, a microblogging network is a complex system when all the key signs of complexity listed above are satisfied. This statement forms the foundation of our research and is key to the construction of a model of microblogging network evolution.

3. Analysis of Empirical Data from Twitter

Empirical data used for our research is a sample of more than 3 million microposts (tweets, retweets, and links) about the first US presidential debates of 2016. The sample includes microposts posted by more than 1 million users from 13:45 on September 26, 2016, to 11:00 on September 27, 2016, with 1-second increments.

Figure 1 shows the total number of microposts vs. time (Twitter time series, MP_t). It is easy to see that MP_t has extraordinary events and unexpected events (bursts).

To estimate the correlation dimension (D_C) and embedding dimension (m), we used the Grassberger–Procaccia algorithm [35]. We obtained $D_C = 3.032$ for $m \leq 6$.

Hence, the process leading to the MP_t series is not random; it depends on a limited number of key parameters [36]. The MP_t series is not stochastic; it is chaotic. For instance, for a stochastic series corresponding to Gaussian noise, $D_C = 9.304$ for $m \leq 13$, and if the series corresponds to generalized Brownian, then noise $D_C = 8.165$ for $m \leq 9$.

Using the R/S analysis, we obtained the Hurst exponent (H). To calculate the fractal dimension of a time series (D_F) we used the algorithm presented in [37]. We obtained the
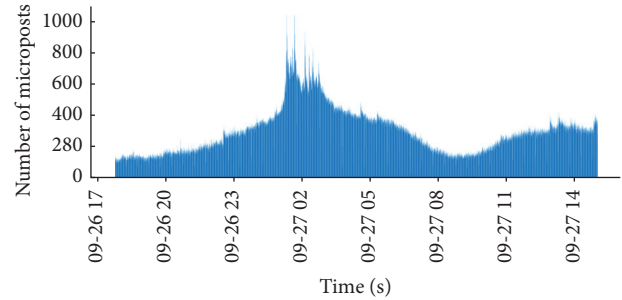
Figure 1: Twitter time series.

Normal distribution
Empirical distribution
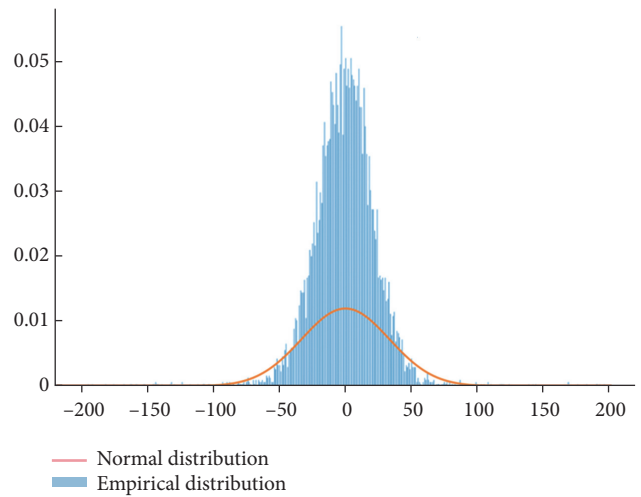
Figure 2: Normal and empirical PDFs.

following results: $H = 0.801$, $D_F = 1.199$. Therefore, MP_t is a fractal time series (the fractal dimension is not an integer and exceeds the topological dimension of the time series). Moreover, MP_t is persistent; i.e. the time series is trend-resistant ($0.5 < H < 1$). Such a time series has a long memory and is inclined to follow trends [38].

Figure 2 shows PDF for the increments (returns) of MP_t time series and PDF for a normal distribution.

Empirical probabilities lie outside the normal PDF in the intervals $(-\infty, -3\sigma)$ and $[3\sigma, +\infty)$. This means that heavy tails exist. D'Agostino's K-squared test [39] also confirms the possibility to reject the null hypothesis about the normality of the distribution at the significance level of 0.01 when the statistics $K^2 = 6419.89$. Another proof that heavy tails exist is presented by the fact that the distribution follows a power law of probability distribution. Figure 3 shows PDF and the complimentary cumulative distribution function (CCDF). Both functions are well approximated by linear functions.

Let us determine the type of noise (parameter β in PSD $S(f) = 1/f^\beta$) for MP_t. To calculate β, we used the detrended fluctuation analysis method (DFA) [40]. After the calculations, we obtained the scaling exponent $\alpha = 1.13$ and the PSD parameter $\beta = 2\alpha - 1 = 1.26$. The β value obtained corresponds rather to a flicker noise ($\beta = 1$) than to any other type of noise. The value $\beta = 1.26$ obtained by the DFA method

FIGURE 3: PDF and CCDF for empirical time series MP_t.

FIGURE 4: PSD for empirical time series MP_t.

FIGURE 5: Autocorrelation function for empirical time series MP_t.

modes in physics). Such a reduction can be justified by the synergetic subordination principle. This principle states that, during the evolution, the hydrodynamic modes suppress the behavior of microscopic degrees of freedom and fully determine the system's self-organization. As a result, the cooperative behavior of a system is determined by several hydrodynamic variables that represent the amplitudes of hydrodynamic modes. This way we do not need an infinite number of microscopic degrees of freedom and there is no need to thoroughly study the microscopic interactions between the users of a social network.

A social network is modeled as a point autonomous dynamical system. This model was chosen because it is possible to compare the results with empirical data provided by the Twitter time series. Each user of a social network can be in one of the two possible states: either passive ($|p\rangle$-state) or active ($|a\rangle$-state). A Twitter user in $|a\rangle$-state can send microposts to other network users. In this state, a network user has enough information to send microposts. If a user is in $|p\rangle$-state (the user does not have enough information), he or she cannot send microposts.

A microblogging social network is an open nonequilibrium system. A social network is capable of information exchange with the environment. The incoming flow of external (for the system) information comes into the system from different sources, for example, from other mass media. This flow feeds the network with information and creates an inverse population of states of network users: $N_{|a\rangle} \gg N_{|p\rangle}$, where $N_{|a\rangle}$ is the number of network users in $|a\rangle$-state, and $N_{|p\rangle}$ is the number of network users in $|p\rangle$-state.

The distribution of Twitter users can be represented with good accuracy by a Boltzmann distribution:

$$N_{|a\rangle} = N_{|p\rangle} \exp\left[-\frac{\left(I_{|a\rangle} - I_{|p\rangle} \right)}{\theta} \right], \qquad (4)$$

where $I_{|a\rangle}$ is the amount of information the users in $|a\rangle$-state possess, $I_{|p\rangle}$ is the amount of information the users in $|p\rangle$-state possess, and θ is a parameter that describes the average intensity of stochastic interactions between the network users. A simple analysis (4) allows us to define two macroscopic network states: a steady state and a nonequilibrium state. If $I_{|a\rangle} - I_{|p\rangle} \gg \theta$, then $N_{|a\rangle} \ll N_{|p\rangle}$. In this case

coincides with the value obtained via the approximation of the time series PSD by a linear function. The PSD obtained by applying fast Fourier transform to MP_t is shown in Figure 4 on a log-log scale. A linear fit yields $\beta = 1.29$.

The autocorrelation function (ACF_τ) for an MP_t time series is described by a decreasing power function (3) with the exponent $\gamma = 0.02$. Hence, this function has long memory.

Figure 5 presents ACF_τ and its linear approximation on a log-log scale. A linear fit gives $\gamma = 0.02$.

4. Microblogging Social Network as a Nonlinear Deterministic Dynamical System

4.1. Main Assumptions for the Model. A social network is a macroscopic system. The number of users for such a system is $\mathcal{N} \gg 1$. This assumption is justified for Twitter, since, according to the existing estimations, $\mathcal{N} \sim 10^8$. In the proposed model, out of all possible degrees of freedom, we choose and consider just a few macroscopic degrees of freedom (phase or dynamic variables corresponding to hydrodynamic

the network is in a steady state. If $I_{|a\rangle} - I_{|p\rangle} \gg \theta$, then $N_{|a\rangle} \gg N_{|p\rangle}$. In this case the network is in a nonequilibrium state. Since a social network is constantly fed with information, it is constantly in a nonequilibrium state, creating an avalanche of microposts. Because of the constant feed of information, a steady state can almost never be reached. It is very important that the existence of chaotic states is a fundamental property of open nonequilibrium systems.

4.2. Phase Variables and Relationships between Them. Let us define the phase variables of a dynamical system. These variables will be used to model Twitter as an open nonequilibrium system. These variables are as follows: $\eta_t \equiv MP_t - MP_0$ is the deviation of the number of microposts (MP_t) from the corresponding equilibrium value (MP_0 is the number of microposts in the steady state); $h_t \equiv I_t - I_0$ is the deviation of aggregated intrasystem information (I_t) the network users possess from the corresponding equilibrium value (I_0 is the aggregated intrasystem information the network users possess when Twitter is in steady state); $S_t \equiv N_{|a\rangle t} - N_{|p\rangle t}$ is the instantaneous difference at the moment of time t between the numbers of strategically oriented social network users (users following a particular strategy) in $|a\rangle$- and $|p\rangle$- states. If $\mathcal{N}$ is the difference between the total number of users in $|a\rangle$- and $|p\rangle$- states, then $\mathcal{N} - S_t$ users act randomly (randomly oriented users).

According to [41], business users and spam users can be considered as strategically oriented users.
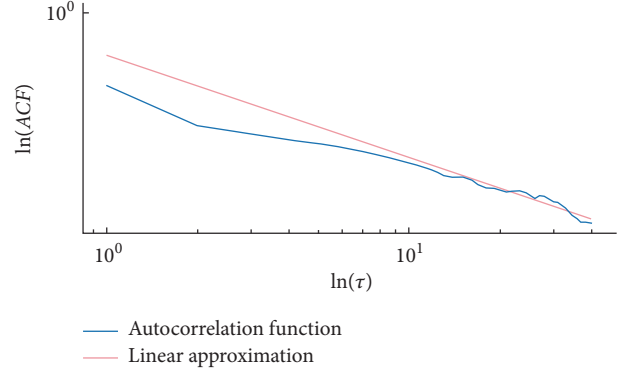
Business users follow marketing and business agendas on Twitter. The profile description strongly depicts their motive, and a similar behavior can be observed in their tweeting behavior. Spammers mostly post malicious tweets at high rates. Automated computer programs (bots) mostly run behind a spam profile and randomly follow users, expecting a few users to follow back.

Personal users and professional users can be considered randomly oriented users. Personal users are casual home users who create their Twitter profile for fun, learning, to get news, etc. These users neither strongly advocate any type of business or product, nor have profiles affiliated with any organization. Generally, they have a personal profile and show a low to mild behavior in their social interaction. Professional users are home users with professional intent on Twitter. They share useful information about specific topics and involve in healthy discussion related to their area of interest and expertise.

Let us determine relationships between the dynamic variables and their rates of change.

The rate of deviation of the number of microposts is determined by the relaxation of a social network into a steady state ($-\eta_t$) and the change in deviation of aggregated intrasystem information from the equilibrium value ($+a_\eta h_t$):

$$\tau_\eta \dot{\eta}_t = -\eta_t + a_\eta h_t. \tag{5}$$

The term $-\eta_t$ in Eq. (5) is due to the relaxation of the social network as a nonequilibrium system. According to Le Chatelier's principle, when a system deviates from the steady state, this generates "forces" that try to restore the system back

to the steady state. As follows from Eq. (5), without the term $a_\eta h_t$ the equation takes the following form:

$$\dot{\eta}_t = -\frac{\eta_t}{\tau_\eta}. \tag{6}$$

The solution to Eq. (6) is given by the function $\eta_t = A\exp(-t/\tau_\eta)$. Hence, $MP_t \longrightarrow MP_0$ when $t \longrightarrow \infty$ (the social network tends to its steady state). In Eq. (6), τ_η is the time of relaxation to the steady state.

The term $a_\eta h_t$ in Eq. (5) can be easily explained: as the deviation of aggregated intrasystem information increases, the rate of the deviation of the number of microposts increases as well.

The rate of the deviation of the aggregated intrasystem information from the equilibrium value is determined by the relaxation of a social network towards a steady state ($-h_t$) and the product $+a_h \eta_t S_t$:

$$\tau_h \dot{h}_t = -h_t + a_h \eta_t S_t. \tag{7}$$

The term $-h_t$ in Eq. (7) is also explained by Le Chatelier's principle, as in Eq. (6). The term $+a_h \eta_t S_t$ appears because the amount of information each user of a social network acquires from a stream of microposts is proportional to the deviation of the number of microposts and depends on the state of the user in the social network. In other words, the average contribution to the deviation of aggregated intrasystem information is proportional to the product of the deviation of the number of microposts and the difference between the numbers of users in $|u\rangle$- and $|l\rangle$- states.

Finally, the third equation describes the change in inversion of population of strategically oriented users and can be written as follows:

$$\tau_S \dot{S}_t = (S_0 - S_t) - a_S \eta_t h_t, \tag{8}$$

where τ_S is the corresponding relaxation time, and S_0 is the initial number of strategically oriented users (this value reflects the intensity of information feeding into the social network). In other words, S_0 is the difference between the numbers of strategically oriented users of a social network which are in $|a\rangle$- and $|p\rangle$-states at the time $t = 0$. The term $-a_S \eta_t h_t$ reflects the effective power that a stream of microposts applies to create aggregated intrasystem information in a social network. This power can be positive or negative.

Thus, the evolution of a microblogging social network can be described by the well-known Lorenz system of equations:

$$\tau_\eta \dot{\eta}_t = -\eta_t + a_\eta h_t$$
$$\tau_h \dot{h}_t = -h_t + a_h \eta_t S_t \tag{9}$$
$$\tau_S \dot{S}_t = (S_0 - S_t) - a_S \eta_t h_t.$$

4.3. Synergetic Interpretation of a Nonlinear Dynamical System. The system of self-consistent equations (9) is a well-known method to describe a self-organizing system. The Lorenz synergetic model was first developed as a simplification of hydrodynamic equations describing the Rayleigh-Bénard heat convection in the atmosphere; it is now a

classical model of chaotic dynamics. Further research on the Lorenz system presented in a series of publications proved that the system provides an appropriate kinetic picture of the cooperative behavior of particles in any macroscopic dynamical system where the actualization of potential order is possible. Processes in such self-organizing complex systems in nonequilibrium state lead to the selection of a small number of parameters from the complete set of variables that describe the system; all other degrees of freedom adjust to correspond to these selected parameters. Following the terminology used in the synergy theory, these parameters are the order parameter (η_t), conjugated field (h_t), and control parameter (S_t). According to the Ruelle-Takens theorem, we can observe a nontrivial self-organization with strange attractors if the number of selected degrees of freedom is three or more.

In the system of equations (9) a_η is a coefficient, and positive constants a_h, a_S are measures of feedback in a social network. Functions η_t/τ_η, h_t/τ_h, $(S_0 - S_t)/\tau_S$ describe the autonomous relaxation of the deviation of the number of microposts, deviations of aggregated information, and inversion of population of strategically oriented users of a social network to the stationary values $\eta_t = 0$, $h_t = 0$, $S_t = S_0$ with relaxation times τ_η, τ_h, τ_S.

Eq. (10) takes into account that, in the autonomous regime, the change in the aforementioned parameters of a social network is dissipative. In addition, Le Chatelier's principle is very important: since the growth of the control parameter S_t is the reason for self-organization, the values η_t and h_t must vary so as to prevent the growth of S_t. Formally, this fact could be explained as the existence of a feedback between the order parameter η_t and the conjugated field h_t. Lastly, a positive feedback between the order parameter η_t and the control parameter S_t leading to the growth of the conjugated field h_t is very important, since this feedback is the reason for self-organization.

4.4. Capabilities and Restrictions of a Deterministic Model for the Interpretation of a Social Network's Complexity. First of all, we have to note that Eq. (9) was first obtained by Edward Lorenz in 1963 as a result of some simplifications of the problem of a liquid layer heated from below. In this problem Eq. (9) is obtained when the flow velocity and temperature of the initial hydrodynamic system are presented as two-dimensional truncated Fourier series and the Boussinesq approximation is used. For the problem of convection in a layer, the Lorenz equations serve as a rough, not very accurate approximation. It is only adequate in the region of regular modes where uniformly rotating convection cells are observed. The chaotic regime typical of Eq. (9) does not describe the turbulent convection. However, the Lorenz equations became a suitable model for describing systems and processes of various natures: convection in a closed loop, single-mode laser, water wheel rotation, financial markets, transportation flows, dissipative oscillator with nonlinear excitation, and some others.

How reliable is the model (9) for the description of the evolution of a microblogging social network? We will consider the model "reliable" if there is a good correlation

between theoretically predicted and empirically observed key signs of complexity of the system. The results of the comparison of key signs of complexity for the theory-based deviations of the number of microposts η_t and the corresponding empirical data are presented below.

As shown earlier (see Eq. (4)), a steady state of the network is almost impossible to achieve due to a constant information feed. Theoretically, a dynamical system (9) has an asymptotically stable zero stationary point as a node for $S_0 = 0$. In this case, $MP_t \approx MP_0$, $I_t \approx I_0$ and $N_{|a\rangle t} \approx N_{|p\rangle t}$ as $t \longrightarrow \infty$. However, in practice, a microblogging network as an open nonequilibrium system always has a non-zero difference between the numbers of strategically oriented users that are in $|a\rangle$-state and in $|p\rangle$-state at the time $t = 0$. Therefore, despite a theoretical feasibility of the steady state for a social network, this state cannot be achieved in practice. When the difference between the numbers of strategically oriented users that are in $|a\rangle$-state and in $|p\rangle$-state at the time $t = 0$ reaches some critical value S_{0c}, Eq. (9) enters a chaotic regime, and a strange attractor appears. A transitional state that corresponds to $S_0 \in (1, S_{0c})$ cannot be realized in practice.

We will consider MP_0 as constant for a long enough period of time and compare different measures for theoretical (η_t, the solution of system (9) in chaotic regime) and empirical data MP_t. The model of a social network presented in the form (9) explains the fractal and chaotic nature of the observed MP_t: $D_C = 2.067$ and $D_F = 1.504$. However, the model (9) cannot explain the observed key signs of complexity of a social network. Theoretical MP_t constitutes a time series without memory (ACF_τ exponentially decreases); PSD is constant (white noise, $\beta = 0$); PDF is multi-modal with "truncated tails" (see Figure 6).

Compactness and multi-modality of the distribution are determined by the existence of three stationary points of the dynamical system (9).

Thus, the Lorenz system (9) is not a reliable model for the description of the evolution of a microblogging social network as a complex system.

5. Microblogging Social Network as a Nonlinear Random Dynamical System

As shown earlier, the nonlinear dynamic model (9) explains the fractality and chaotic nature of empirical MP_t as well as the dissipative nature of the system. On the other hand, Eq. (9) cannot explain some other phenomena found in empirical data, and first of all, the key signs of complexity of a social network: a power law of PDF, $1/f$-noise, and long memory. Let us consider different ways of improving (generalizing) Eq. (9) in order to adequately describe a microblogging social network.

Since the correlation dimension and embedding dimension of the empirical time series ($D_C = 3.032$, $m \leq 6$) exceed the corresponding theoretical values ($D_C = 2.067$, $m \leq 4$), one of the ways to improve Eq. (9) is to increase the number of phase variables of the dynamic system. Another approach to improving Eq. (9) is to consider the self-consistent behavior of the order parameter, conjugated
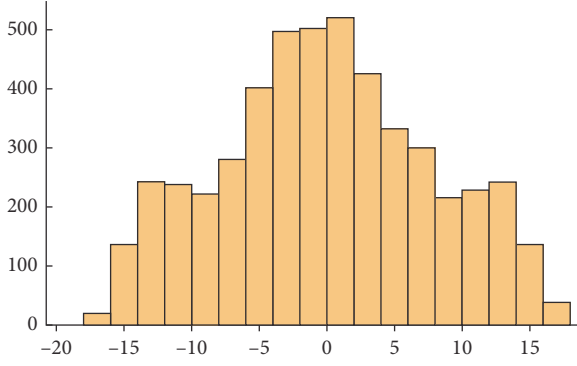
Figure 6: Histogram for the time series MP_t.

field, and control parameter taking into account the noise for each of those parameters. Different generalizations of Eq. (9) have been proposed and studied by Alexander Olemsky and collaborators [42–44], in particular, in the context of its applications to the study of self-organization of continuum, evolution of financial markets and economical structure of society, cooperative behavior of active particles, and self-organized criticality.

Taking into account stochastic terms and the fractionality of the order parameter, Eq. (9) takes the following form:

$$\tau_\eta \dot{\eta}_t = -\eta_t^\alpha + a_\eta h_t + \sqrt{I_\eta} \xi_t$$

$$\tau_h \dot{h}_t = -h_t + a_h \eta_t^\alpha S_t + \sqrt{I_h} \xi_t \qquad (10)$$

$$\tau_S \dot{S}_t = (S_0 - S_t) - a_S \eta_t^\alpha h_t + \sqrt{I_S} \xi_t.$$

In Eq. (10), I_i are noise intensities for each phase variable; ξ_t is white noise, where $\langle \xi_t \rangle = 0$, $\langle \xi_t \xi_{t\prime} \rangle = \delta(t - t\prime)$; $\alpha \in (0, 1]$. The random dynamic system (RDS) (10) is a generalization of the deterministic dynamic system (9) where stochastic sources are added, the feedback is weakened, and the order parameter is relaxed. The replacement of the order parameter η_t by a smaller value η_t^α ($\alpha \leq 1$) means that the process of ordering influences the self-consistent behavior of the system to a lesser extent than it does in the ideal case of $\alpha = 1$.

For the convenience of the analysis of Eq. (10) we will transform it into a dimensionless form. Then time t, deviation of the number of microposts (η_t), deviation of the aggregated intrasystem information (h_t), the difference between the numbers of strategically oriented users in different states (S_t), and corresponding noise intensities (I_i) will be scaled as follows:

$$t_c \equiv \tau_\eta (a_\eta a_h)^{(\alpha-1)/(2\alpha)},$$

$$h_c \equiv (a_\eta^2 a_h a_S)^{-1/2},$$

$$\eta_c \equiv (a_h a_S)^{-1/(2\alpha)},$$

$$S_c \equiv (a_\eta a_h)^{-1/2},$$

$$I_\eta^c \equiv (a_h a_S)^{-1/\alpha},$$

$$I_h^c \equiv (a_\eta^2 a_h a_S)^{-1},$$

$$I_S^c \equiv (a_\eta a_h)^{-1/2}. \qquad (11)$$

Now Eq. (10) can be written down as follows:

$$\dot{\eta} = -\eta^\alpha + h + \sqrt{I_\eta} \xi$$

$$\frac{\tau_h}{t_c} \dot{h} = -h + \eta^\alpha S + \sqrt{I_h} \xi \qquad (12)$$

$$\frac{\tau_S}{t_c} \dot{S} = (S_0 - S) - \eta^\alpha h + \sqrt{I_S} \xi.$$

Let us analyze RDS (12) in adiabatic approximation when the characteristic relaxation time of the number of microposts in a network considerably exceeds the corresponding relaxation times of aggregated intrasystem information and the number of strategically oriented users: $\tau_\eta \gg \tau_h, \tau_S$. This means that aggregated intrasystem information $h \approx h(\eta)$ and the number of strategically oriented users $S \approx S(\eta)$ follow the variation in the deviation of the number of microposts (η). When $\tau_\eta \gg \tau_h, \tau_S$, the subordination principle allows us to set $(\tau_h/t_c)\dot{h} = (\tau_S/t_c)\dot{S} = 0$ in Eq. (12), i.e., to disregard the fluctuations in $h \approx h(\eta)$ and $S \approx S(\eta)$.

For a microblogging social network functioning as an open nonequilibrium system, the adiabatic approximation means that when the external information feed tends to zero, the stream of microposts slowly decreases and at the same time the aggregated intrasystem information and the number of strategically oriented users in active state decrease as well.

An adiabatic approximation is a necessary condition for the transformation of the three-dimensional RDS with additive noise (12) into a one-dimensional RDS with multiplicative noise of the following form:

$$\tau_\eta \dot{\eta} = f(\eta) + \sqrt{I(\eta)} \xi. \qquad (13)$$

The terms of Eq. (13) corresponding to the drift and diffusion (intensity of the chaotic source) have the following form:

$$f(\eta) \equiv -\eta^\alpha + S_0 \eta^\alpha (1 + \eta^{2\alpha}), \qquad (14)$$

$$I(\eta) \equiv I_\eta + (I_h + I_S \eta^{2\alpha})(1 + \eta^{2\alpha})^2. \qquad (15)$$

The Langevin equation (13) has an infinite set of random solutions η. Their probability distribution ($p(\eta, t)$) is given by the Fokker-Planck equation:

$$\partial_t p(\eta, t) = \partial_\eta \left[-f(\eta) p(\eta, t) + \partial_\eta (I(\eta) p(\eta, t)) \right]. \qquad (16)$$

In the stationary case ($\partial_t p(\eta, t) = 0$) the distribution $p(\eta, t)$ is given by the following relationship:

$$p(\eta) \propto I^{-1}(\eta) \exp \left[\int \frac{f(\eta)}{I(\eta)} d\eta \right]. \qquad (17)$$

As a result, the stationary probability distribution density of the deviation of the number of microposts from the corresponding equilibrium value has the following form:

$$p(\eta) = Z^{-1}\left(1 + \eta^{2\alpha}\right)^{-2}\exp\left[\int \frac{\left(1 + \eta^{2\alpha}\right)^{-2}}{\eta^{\alpha}}d\eta\right], \quad (18)$$

where Z is a normalization constant.

Before we draw any conclusion about distribution (18), let us direct our attention to one significant fact that distinguishes theory from practice. Distributions of real systems and processes regardless of their nature cannot have an infinite expected value or variance. Therefore, power-law PDFs like $p(x) \propto x^{-2\alpha}$ (2α is chosen for the purposes of analysis of expression (18)) are approximate and not valid for large x. The exponential decrease of PDF corresponds to the intermediate asymptotics, and in practice instead of heavy tails we should have semi-heavy tails (see distribution in Figure 2):

$$p(\eta) \propto \eta^{-2\alpha}\mathscr{P}\left(\frac{\eta}{\eta_c}\right), \quad (19)$$

where the scaling function $\mathscr{P}(\eta/\eta_c)$ is approximately constant at $\eta \cong \eta_c$ and quickly decreases when $\eta \longrightarrow \infty$. Here the "heaviness of the tail" is shifted toward the intermediate range of η values. Thus, the dimensionless deviation of the number of microposts η scaled for η_c serves as a scaling variable η/η_c in (19). Since the integral in Eq. (16) is regular at $\eta \longrightarrow 0$, the PDF obtained has a power-law form.

The power law for PDF of the deviation of the number of microposts η_t, which is equivalent to MP_t for large times, was obtained and justified analytically. However, we could not obtain analytical expressions for PSD, ACF_τ, or the correlation and fractal dimensions. Therefore, we present below the results of numerical calculations for a family of realizations of RDS (13) for $\alpha = 0.5$ based on algorithms studied and used earlier.

Let us determine the type of noise typical for η_t. We used the DFA method to calculate β. We obtained the scaling exponent $\alpha = 1.18$ and $\beta = 2\alpha - 1 = 1.36$. The value obtained for β corresponds rather to flicker noise ($\beta = 1$) than to any other type of noise. The value $\beta = 1.18$ obtained by the DFA method is close to the value obtained through fitting PSD time series by a linear function. PSD obtained by applying fast Fourier transform to η_t is presented in Figure 7 in log-log scale. A linear fit gives $\beta = 1.34$.

ACF_τ for the time series η_t decreases by following the power law (3) with the exponent $\gamma = 0.04$ and hence has double memory. Figure 8 shows ACF_τ for η_t in log-log scale.

To estimate the correlation dimension (D_C) and embedding dimension (m), we used the Grassberger–Procaccia algorithm. We obtained $D_C = 2.852$ for $m \leq 5$. Hence, the process that leads to the series η_t is not random; it is controlled by a limited number of key parameters. The series η_t is chaotic rather than stochastic.

Using the results of R/S analysis we determined the Hurst exponent (H). To calculate the fractal dimension of the time series (D_F) we used the algorithm described in [20]. We
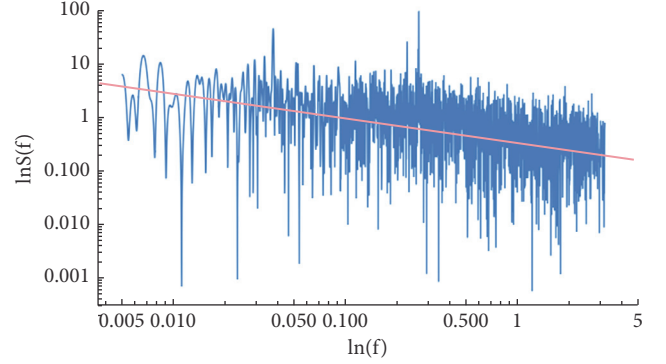
FIGURE 7: PSD for η_t time series.
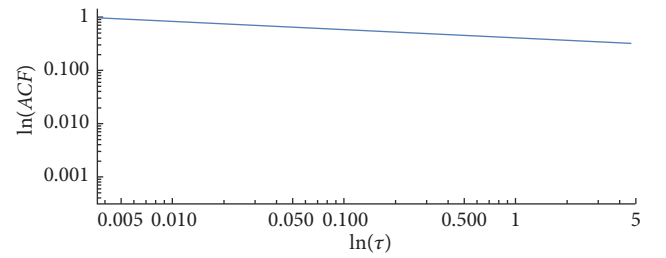
FIGURE 8: Autocorrelation function for η_t time series.

obtained $H = 0.765$, $D_F = 1.235$. Hence, η_t is a persistent fractal time series. Such time series has a long-term memory and tends to follow trends.

Therefore, the generalized Lorenz system (12) adequately models the evolution of a microblogging social network as a complex system. The characteristics of η_t-realizations of RDS (12) are quantitatively close to the corresponding characteristics of empirical time series.

6. Results and Discussion

For the convenience of further discussion, Table 1 presents the results of calculations of key characteristics and properties of complex systems (i.e., systems that tend to have unexpected and/or extraordinary events).

The empirical time series of microposts has all the key properties of complexity: a power-law PDF, noise that is close to flicker noise, time correlations with long memory, and scale invariance in a time series of microposts. The existence of bursts in time series of microposts (see Figure 1) allows us to conclude that a microblogging network is a complex system, and it is far from equilibrium. The time series of microposts is characterized by scale invariance; i.e., it is a fractal time series. Such time series, in particular, are characterized by power-law PDFs caused by an avalanche-like increase of the number of microposts (see bursts in Figure 1) after a "chain reaction" of reposting. An avalanche-like increase of the number of microposts is possible if a user coordinates his actions with his followers, followers of those followers, and so on. This defines a connection between the catastrophic behavior and integrity of a microblogging network.

TABLE 1: Characteristics of empirical and theoretical time series.

Time Series	PDF	β	γ	D_C	m	D_F	H
Empirical	Power Law	1.29	0.02	3.032	6	1.199	0.801
Lorenz	Compact	0	Exponential	2.067	4	1.504	0.496
Generalized Lorenz	Power Law	1.36	0.04	2.852	5	1.235	0.765

For a description of the evolution of a microblogging network, the nonlinear dynamical system model (9) is a rough, not very accurate approximation. First, the model does not predict the occurrence of catastrophic values in a time series of the number of microposts which would signify the complexity of a microblogging network, or the existence of long memory or the time series' tendency to follow trends. Despite this deficiency, Eq. (9) allows one to study social networks far from equilibrium (see distribution (4) and comments thereon), and it also explains the existence of dynamical chaos in a time series as well as their fractality.

The nonlinear random dynamical system (10) is a generalization of the model (9) that accounts for external stochastic sources and the fractionality of the order parameter (weakening of feedback and relaxation of the order parameter). This model adequately describes the evolution of a microblogging system.

Quantitative characteristics of the model (10) in adiabatic approximation are close to the corresponding characteristics of the empirical time series of microposts (see Table 1). An adiabatic approximation allows us to reduce a three-dimensional random adiabatic system with additive noise (10) to a one-dimensional random dynamical system with exponential multiplicative noise (13).

7. Conclusions

The main results of this research were obtained by analyzing a single time series of microposts whose values however constitute a representative sample. Similar results of analysis of an empirical time series of a microblogging network are presented in [24–29]. We cannot claim that the time series samples studied by us or other researchers are representative, which would be essential for a generalization of the results onto the entire general population. In the framework of this approach, it is necessary to analyze all the available data on microposts and users collected since the beginning of the microblogging network. However, this step could be avoided if we consider the scale invariance of social networks. This allows us to extrapolate and interpolate the results of the network analysis onto any large or small scale. Hence, the fractality of a sample predetermines the fractality of the entire network. A justification of the scale invariance for Twitter is presented in [45]. Therefore, the conclusion about the complexity of microblogging networks in the framework of the paradigm of complexity is justified.

What follows from the fact that a microblogging network is complex? We can give two answers to this question. The first is connected with the possibility of second-order phase transitions in a microblogging network; the second concerns the analysis and prediction of a time series of microposts. Let us elaborate on each answer.

It has been established that time series of microposts are characterized by long-range time correlations. This is true both for empirical time series and for realizations of the random dynamical system (13). Long-range correlations and other characteristics of time series discussed above are typical of critical phenomena such as second order phase transitions.

For simplicity, let us consider the kinetics of a phase transition in a microblogging network in the framework of the model (9) taking into account the stochasticity of the feed (the difference in the initial number of strategically oriented users in active and passive states S_0). In this case, it can be shown (a detailed proof lies outside the scope of this paper) that as S_0 increases and exceeds a certain critical value, a microblogging network evolves according to the strategy chosen by a relatively small number of strategically oriented users. The aforementioned avalanche-like increase of the number of microposts takes place. The critical value of S_0 is determined by the geometric mean of the total and critical values of the number of strategically oriented users. A formalism that leads to the above result is presented in [42].

The results obtained in this paper are valuable from both theoretical and practical points of view. Firstly, they show that the systems under consideration (in this case the number of microposts) are deterministic despite having noise components (i.e., they are not stochastic). This allows us to use the theory of dynamical systems and analyze the time series of microposts in a different way, using the dimension theory and the theory of dynamic systems. Secondly, the values of invariants obtained can help solve the problem of prediction. For example, the embedding dimension shows how many terms of a time series determine the next term, whereas the correlation entropy and the largest Lyapunov exponent allow us to estimate the time of predictability of the system.

In conclusion, we would like to mention that there exist many interesting problems that are not studied yet, such as critical phenomena of self-organization in microblogging networks based on the analysis of the nonlinear random dynamic system (10). This will be the subject of our future research.

Data Availability

Data was obtained by hydrating a list containing 3,183,202 identifiers of tweets from the set of 12 lists of identifiers provided by Harvard University. This list is about the 2016

USA presidential elections: «2016 United States Presidential Election Tweet Ids» (2016). The list was created by Justin Littman, Laura Wrubel, and Daniel Kerchner. The authors of the list used SocialFeed service to gather data after the first debates. Tweets on the subsequent debates were not included in the sample. The sample obtained has about 1 million empty entries. This happened because some users whose identifiers were in the initial list later removed their tweets or made them private. The resulting sample has the following characteristics: a micropost can correspond to one hashtag or several hashtags (#debate, #debates, #debatenight, #debate2016, #debates2016); the presence of the micropost's author in the list of followers of one or several users (CPD (@debates), Hillary Clinton (@HillaryClinton) и Donald J. Trump (@realDonaldTrump)); 2,290,855 microposts; 934,656 users; 76,458 time intervals; one-second increments. After the list of tweets was received, the information was extracted in id:original_id format. Here id is a unique identifier of the user who made the retweet; the original_id is a unique identifier of the user who made the initial tweet. If a tweet is not a retweet, id and original_id coincide.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] D. J. De Solla Price, "Networks of scientific papers," *Science*, vol. 149, no. 3683, pp. 510–515, 1965.

[2] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[3] R. Albert and A. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.

[4] C. T. Butts, "The complexity of social networks: Theoretical and empirical findings," *Social Networks*, vol. 23, no. 1, pp. 31–71, 2001.

[5] J. Skvoretz, "Complexity theory and models for social networks," *Complexity*, vol. 8, no. 1, pp. 47–55 (2003), 2002.

[6] M. G. Everett, "Role similarity and complexity in social networks," *Social Networks. An International Journal of Structural Analysis*, vol. 7, no. 4, pp. 353–359, 1985.

[7] H. Ebel, J. Davidsen, and S. Bornholdt, "Dynamics of social networks," *Complexity*, vol. 8, no. 2, pp. 24–27 (2003), 2002.

[8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. W. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[9] S. Tabassum, F. S. F. Pereira, S. Fernandes, and J. Gama, "Social Networks Analysis: An Overview," *WIREs Data Mining and Knowledge Discovery*, pp. 1–21, 2018.

[10] S. Saganowski, B. Gliwa, P. Bródka, A. Zygmunt, P. Kazienko, and J. Kozlak, "Predicting community evolution in social networks," *Entropy*, vol. 17, no. 5, pp. 3053–3096, 2015.

[11] P. De Meo, F. Messina, D. Rosaci, and G. M. L. Sarné, "Forming time-stable homogeneous groups into Online Social Networks," *Information Sciences*, vol. 414, pp. 117–132, 2017.

[12] A. Grabowski and R. A. Kosiński, "Ising-based model of opinion formation in a complex network of interpersonal interactions," *Physica A: Statistical Mechanics and its Applications*, vol. 361, no. 2, pp. 651–664, 2006.

[13] S. Dasgupta, R. K. Pan, and S. Sinha, "Phase of Ising spins on modular networks analogous to social polarization," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 80, no. 2, Article ID 025101, 2009.

[14] G. Bianconi, "Mean Field Solution of the Ising Model on a Barabási-Albert Network," *Physics Letters A*, vol. 303, no. 2-3, pp. 166–168, 2002.

[15] C. Li, F. Liu, and P. Li, "Ising model of user behavior decision in network rumor propagation," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 5207475, 2018.

[16] J.-L. Guo, Q. Suo, A.-Z. Shen, and J. Forrest, "The evolution of hyperedge cardinalities and bose-Einstein condensation in hypernetworks," *Scientific Reports*, vol. 6, Article ID 33651, 2016.

[17] M. Faccin, T. Johnson, J. Biamonte, S. Kais, and P. Migdał, "Degree Distribution in Quantum Walks on Complex Networks," *Physical Review X*, vol. 3, no. 4, Article ID 041007, 2013.

[18] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 1, Article ID 016110, 2006.

[19] B. P. Chamberlain, J. Levy-Kramer, C. Humby, and M. P. Deisenroth, "Real-time community detection in full social networks on a laptop," *PLoS ONE*, vol. 13, no. 1, Article ID e0188702, 2018.

[20] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos, "Nonlinear dynamics of information diffusion in social networks," *ACM Transactions on the Web (TWEB)*, vol. 11, Article 11, no. 2, 2017.

[21] N. Hegde, L. Massoulie, and L. Viennot, "Self-organizing flows in social networks," *Theoretical Computer Science*, vol. 584, no. 13, pp. 3–18, 2015.

[22] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized Criticality: An Explanation of 1/f-noise," *Physical Review Letters*, vol. 59, no. 4, pp. 381–384, 1987.

[23] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized Criticality," *Physical Review A*, vol. 38, no. 1, pp. 364–374, 1988.

[24] P. Bak, *How Nature Works: The Science of Self-organized Criticality*, Springer-Verlag, 1996.

[25] B. Tadić, M. M. Dankulov, and R. Melnik, "Mechanisms of self-organized criticality in social processes of knowledge creation," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 96, no. 3, Article ID 032307, 2017.

[26] M. Aguilera, I. Morer, X. Barandiaran, and M. Bedia, "Quantifying Political Self-Organization in Social Media. Fractal patterns in the Spanish 15M movement on Twitter," in *Proceedings of the 12th European Conference on Artificial Life*, pp. 395–402, Michigan, USA, 2013.

[27] K. Lyudmyla, B. Vitalii, and R. Tamara, "Fractal time series analysis of social network activities," in *Proceedings of the 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*, pp. 456–459, IEEE, Kharkov, Ukraine, October 2017.

[28] T. De Bie, J. Lijffijt, C. Mesnage, and R. Santos-Rodriguez, "Detecting trends in twitter time series," in *Proceedings of the 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, Vietri sul Mare, Salerno, Italy, September 2016.

[29] A. Mollgaard and J. Mathiesen, "Emergent user behavior on twitter modelled by a stochastic differential equation," *PLoS ONE*, vol. 10, no. 5, pp. 1–12, 2015.

[30] A. Dmitriev, V. Dmitriev, O. Tsukanova, and S. Maltseva, "A nonlinear dynamical approach to the interpretation of microblogging network complexity," *Studies in Computational Intelligence*, vol. 689, pp. 390–400, 2017.

[31] B. Tadić, V. Gligorijević, M. Mitrović, and M. Šuvakov, "Co-evolutionary mechanisms of emotional bursts in online social dynamics and networks," *Entropy*, vol. 15, no. 12, pp. 5084–5120, 2013.

[32] M. M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos*, Touchstone, New York, USA, 1993.

[33] O. A. Tsukanova, E. P. Vishnyakova, and S. V. Maltseva, "Model-based monitoring and analysis of the network community dynamics in a textured state space," in *Proceedings of the 16th IEEE Conference on Business Informatics, CBI 2014*, pp. 44–49, Switzerland, July 2014.

[34] Ming Li, "Fractal Time Series—A Tutorial Review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.

[35] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189–208, 1983.

[36] M. Z. Ding, C. Grebogi, E. Ott, T. Sauer, and J. A. Yorke, "Estimating correlation dimension from a chaotic time series: when does plateau onset occur?" *Physica D: Nonlinear Phenomena*, vol. 69, no. 3-4, pp. 404–424, 1993.

[37] M. M. Dubovikov, N. V. Starchenko, and M. S. Dubovikov, "Dimension of the minimal cover and fractal analysis of time series," *Physica A: Statistical Mechanics and its Applications*, vol. 339, no. 3-4, pp. 591–608, 2004.

[38] B. B. Mandelbrot and J. W. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM*, vol. 10, no. 4, pp. 422–437, 1968.

[39] R. B. D'Agostino, A. Belanger, and R. B. D'Agostino, "A suggestion for using powerful and informative tests of normality," *The American Statistician*, vol. 44, no. 4, pp. 316–321, 1990.

[40] R. B. Govindan, J. D. Wilson, H. Preil, H. Eswaran, J. Q. Campbell, and C. L. Lowery, "Detrended fluctuation analysis of short datasets: an application to fetal cardiac data," *Physica D: Nonlinear Phenomena*, vol. 226, no. 1, pp. 23–31, 2007.

[41] M. M. Uddin, M. Imran, and H. Sajjad, "Understanding Types of Users on Twitter," in *Proceedings of 6th ASE International Conference in Social Computing*, Stanford, USA, 2014.

[42] A. I. Olemskoi, A. V. Khomenko, and D. O. Kharchenko, "Self-organized criticality within fractional Lorenz scheme," *Physica A: Statistical Mechanics and its Applications*, vol. 323, no. 1-4, pp. 263–293, 2003.

[43] A. I. Olemskoĭ, "Theory of stochastic systems with singular multiplicative noise," *Physics-Uspekhi*, vol. 41, no. 3, pp. 269–301, 1998.

[44] A. I. Olemskoi and A. V. Khomenko, "Three-Parameter Kinetics of a Phase Transition," *Journal of Theoretical and Experimental Physics*, vol. 81, no. 6, pp. 1180–1192, 1996.

[45] S. Aparicio, J. Villazón-Terrazas, and G. Álvarez, "A Model for Scale-Free Networks: Application to Twitter," *Entropy*, vol. 17, no. 12, pp. 5848–5867, 2015.

Research Article

Simulation of Knowledge Transfer Process Model Between Universities: A Perspective of Cluster Innovation Network

Fang Wei [ID] **and Xiao Limin** [ID]

School of Management, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Fang Wei; fwx1998@nwpu.edu.cn

Combined with the basic properties of the cluster innovation network, with the cluster innovation network, which can be composed of different universities that have knowledge potential difference as the research object, the knowledge transfer process is divided into four stages: knowledge externalization, knowledge sharing, knowledge innovation, and knowledge internalization, and the article constructs a knowledge transfer process model through introducing explicit knowledge and tacit knowledge conversion effect mechanism. According to the theory of complex adaptive system, the principle of network connection oriented the knowledge potential difference and the characteristic of the explicit knowledge and tacit knowledge within universities. We research the knowledge transfer process of universities using the system simulation method and focus on the evolution mechanism of the cluster innovation network's knowledge level at knowledge externalization and knowledge sharing stage. It further reveals the basic topology structure and dynamic evolution law of universities cluster innovation network. We find that both knowledge externalization efficiency and knowledge learning ability have positive correlation with the general knowledge level of network. The evident small-world network characteristic emerges during the dynamic evolution of universities cluster innovation network. Meanwhile, there exists a coupling evolution between the knowledge level of universities and the topology structure of the cluster innovation network.

1. Introduction

In order to enhance the comprehensive strength and international competitiveness of higher education, China proposed the goal of "Double First-Rate" construction with world-class universities and world-class disciplines from a strategic perspective in 2015. Strengthening the capacity for independent innovation and further increasing the iconic innovation achievements with significant influence at home and abroad have become a crucial way to achieve the goal of "Double First-Rate." Studies have shown that industrial clusters can positively enhance knowledge dissemination and innovation performance [1, 2]; integrating resources through network relationships can improve innovation performance [3]. Therefore, universities within the cluster innovation network through cooperation can obtain important innovation resources to stimulate innovation vitality and improve their level of knowledge.

In the era of knowledge-based economy, as the environment changes and the complexity of innovation deepens, it is difficult for individual innovation to meet innovative demands. At this time, cooperative innovation under network conditions is becoming more and more popular [4]. Meanwhile, the innovation process shows characteristics of complex knowledge network [5]. Thus, the knowledge network connected by knowledge subjects such as universities, enterprises, and scientific research institutions has become the core platform of innovation activities. Knowledge subjects integrate resources and cooperate deeply through establishing formal and informal relationships to acquire and share knowledge and information resources embedded in their internal and external networks, and ultimately achieve the purpose of creating new knowledge [6]. With the implementation of the innovation-driven strategy, the cluster innovation network has become a new model and mechanism for dealing with innovation. Strong cluster collaboration can

enhance innovation capability and allow organizations to achieve their goal that could not be achieved alone [7]. As an important carrier of knowledge flow, the cluster innovation network is a self-organizing emergence in which internal and external innovation subjects of the cluster innovation network adapt to the complexity of innovation [8]. The partner selection behavior of subjects influences the evolution of the innovation network structure [9]. Network structure is a crucial factor which influences the knowledge transfer and innovation performance [10–12]. Therefore, there is a complex relationship between the knowledge transfer and the network structure in the innovation network and many studies have discussed this superficially and deeply.

Many studies showed the surface relationships between knowledge transfer and network structure: IM Taplin [13] studied network structure and knowledge transfer in cluster evolution using the methods of qualitative analysis; Fritsch M et al. [14] focused on knowledge transfer in a sample of 16 German regional innovation networks with almost 300 firms and research organizations involved and found that strong ties are more beneficial for knowledge exchange than weak ties through the case study; Kim and Park [15] constructed the knowledge diffusion process of R&D network to investigate the impact of network structure on the performance of knowledge diffusion; the results show that the small-world network is the most efficient and equitable structure toward effective knowledge diffusion. In addition, many scholars were aware of the complex network and adaptive system characteristics of multi-agent cluster cooperative networks. They analyzed the deep-rooted mechanism of knowledge transfer process and dynamic evolution law in the innovation network using modern multi-intelligent simulation methods; B He and G Song [16] established the differential dynamic model of tacit knowledge transfer efficiency and made example simulation to research how cluster network structure feature influences tacit knowledge transfer process; Wang [17] constructed the knowledge transfer diffusion process model of the cluster innovation cooperation network and analyzed the impact of individual motivation on knowledge transfer and diffusion performance using the intelligent simulation method; MA Xuejun et al. [18] built the industry alliances knowledge transfer network model with the quantitative analysis of the simulation example from a complex network perspective.

Although most studies have analyzed the relationships between knowledge transfer and network structure among enterprises within cluster innovation network from various aspects, studies on the factors affecting knowledge transfer and network evolution process in universities-oriented cluster networks are scarce. In addition, a large majority of researches take the abstract and general knowledge as the research object; they do not divide the research object into explicit and tacit knowledge and neglect the transformation influence mechanism of explicit and tacit knowledge in the process of knowledge transfer. Moreover, the above lack quantitative researches on each phase of knowledge transfer.

Therefore, based on the previous researches on knowledge network and knowledge transfer process, this study establishes the model of knowledge transfer process of universities from the perspective of the cluster innovation network and explores quantitatively conversion influence mechanism of explicit and tacit knowledge from the attribute dimension of knowledge (explicit and tacit knowledge) using complex adaptive system theory and system simulation method. Additionally, this paper focuses on the relationship between the mechanism of partner selection based on the knowledge potential difference of knowledge subjects, knowledge level, and the cluster innovation network structure in the knowledge sharing stage. Moreover, the basic topology structure and dynamic evolution law of universities innovation network can be revealed.

2. Theoretical Framework

After the concept of knowledge transfer was first proposed by Teece [19] in 1977, many scholars at home and abroad have proposed different knowledge transfer models through researching and exploring. The most representative is the SECI model presented by Nonaka and Takeuchi [20]. They first combine the knowledge attribute dimension (explicit and tacit knowledge) with knowledge transfer and propose the organizational knowledge creation spiral which divide knowledge transfer into socialization, externalization, combination, and internalization. The nature of this model is spiral structure in the process of self-transformation and mutual transformation of explicit and tacit knowledge.

At the same time, the cluster innovation network has gradually become an important platform and support for knowledge transfer and knowledge innovation among the various knowledge subjects in the cluster. Industrial cluster networks are the context of knowledge transfer between different subjects [21], so the essence of the cluster innovation network with universal characteristics of the obvious knowledge network is the knowledge network. The network structure can present complex evolutionary dynamics when knowledge transfer is within the innovation network. Meanwhile, some studies have shown that the innovation results formed by knowledge transfer in the cluster network are often larger than the results of the individual innovation [22–24]. Thus, the knowledge transfer within the cluster innovation network is a key link of cluster innovation activities. It is an important factor of the competitiveness of cluster firms, innovation, and development of industrial clusters [21]. In the whole process of transfer, the knowledge subjects in the network not only enhance knowledge levels, but also change the breadth and depth of knowledge stock [25] (knowledge potential difference) through learning, transformation, and accumulation of explicit and implicit knowledge. Marjolein CJ [26] points out that knowledge transfer behaviors are difficult to occur when knowledge potential differences of diverse subjects are too large or too small. Therefore, the changed knowledge potential differences can in turn affect cooperation relationships among subjects, thus leading to the evolution of the cluster innovation network and its topology [27], and the cluster innovation network has a significant small-world phenomenon in the process of dynamic evolution.
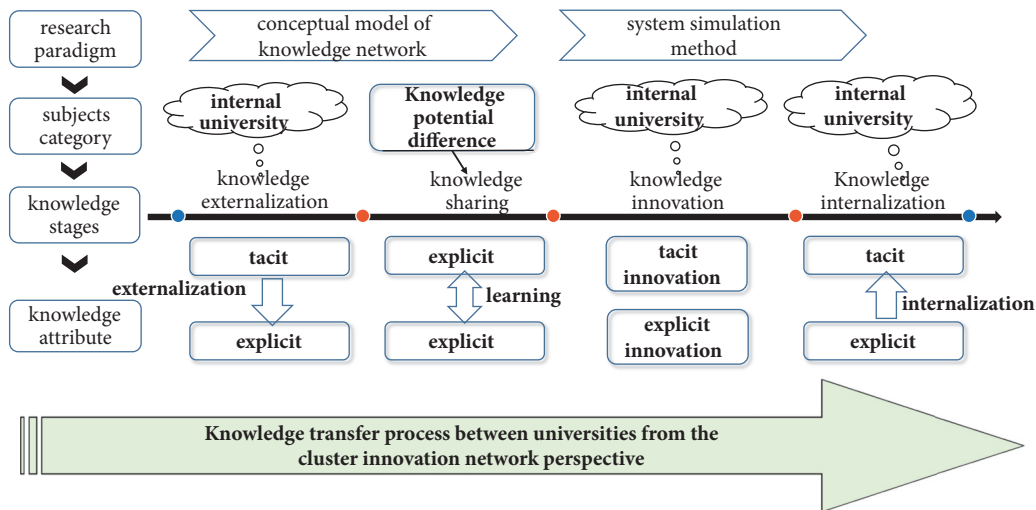
FIGURE 1: Universities' knowledge transfer process model from the cluster innovation network perspective.

In the cluster innovation network, the process of knowledge transfer between subjects is not only an exchange of knowledge, but also a knowledge innovation and its spiral growth. However, knowledge transfer, knowledge innovation, and knowledge growth are closely related to the characteristics of innovation subjects. Compared with firms, universities, as a special knowledge-intensive organization, have more comprehensive knowledge, more diverse levels, and more prospective research fields. Universities have large pieces of explicit knowledge, as academic results, research data, and so on; they, especially, have formed massive tacit knowledge over a long period of time, such as campus culture, training methods, research methods, and thinking patterns of scholars or students [28]. Overall, it is generally considered that universities have unique characteristics of explicit and tacit knowledge: their tacit knowledge is a lot richer than explicit knowledge and they have low level of externalization of tacit knowledge.

This paper draws on the SECI model and combines the network connection principle of knowledge potential difference with the unique explicit and tacit knowledge characteristics of universities to establish the four-stage model of knowledge transfer from the perspective of universities' cluster innovation network (Figure 1).

Knowledge externalization is the first stage of knowledge transfer, which is realized within knowledge subjects. It mainly converts noncoding tacit knowledge into the explicit knowledge expressed by words, graphs, formulas, and so on through coding and simulation. This stage plays a vital role in the process of knowledge transfer. Meanwhile, it is imperative to make tacit knowledge externalize before knowledge sharing stage [29], because knowledge sharing requires the necessary communication and mutual cooperation among knowledge subjects, but tacit knowledge with highly tacitness and its characteristics make it hard to be shared. The externalization of tacit knowledge can promote knowledge flow and improve knowledge transfer performance [30].

In the stage of knowledge sharing, this study assumes that only explicit knowledge can be exchanged between knowledge subjects based on the characteristics of tacit knowledge, such as tacitness, contingency, and difficulty to circulate. After the first stage of tacit knowledge externalization, it has eliminated obstacles that tacit knowledge is difficult to flow to a certain extent. When the knowledge potential difference is within appropriate range, the knowledge subjects (universities) in the network establish a learning cooperation relationship to exchange and learn explicit knowledge. As knowledge exchange constantly goes deeper, both the knowledge levels and the knowledge similarity of subjects are getting higher and higher. According to Marjolein CJ [26], if the innovation subjects' knowledge levels are too similar or too different than each other, the cooperation will be unnecessary in the innovation cluster. At this time, the knowledge gaps between each other are getting smaller or wider until cooperation conditions are not met. In order to break through the current network and further enhance the level of knowledge, some knowledge subjects will seek new partners, which will further arouse continuous evolution of the cluster innovation network. At the same time, there will be a small-world phenomenon in process of evolution: highly clustered and small characteristic path length [31].

As the third stage of knowledge transfer, the major task of knowledge innovation is that each subject analyzes the knowledge learned from the knowledge sharing stage and interacts with its own the existing knowledge. The result of different knowledge interaction can lead to knowledge innovation [20]. Knowledge innovation is based on knowledge sharing; the new knowledge learned by various knowledge subjects will have an impact on the explicit knowledge and tacit knowledge they had before. So at this stage, explicit knowledge and tacit knowledge will innovate, thereby changing the overall level and stock of knowledge.

Knowledge internalization, as the last stage, is not only a value transformation and formation stage of the knowledge

transfer process, but also a knowledge promotion and application stage. It can be regarded as the reverse behavior of externalization of tacit knowledge in form: implicitization of explicit knowledge on the basis of innovation. In this stage, knowledge subjects absorb and digest explicit knowledge and internalize it into a higher level of tacit knowledge to achieve mastery and sublimation of knowledge.

3. Model Construction

3.1. Introduction to the Model. According to Valk [32] and Hermans F [33] et al., the cluster innovation network consists of nodes, which represent knowledge subjects (universities), and links, which represent the relationships of knowledge exchange and cooperation between universities. This research assumes that the number of knowledge subjects (nodes) in the cluster innovation network is N and the initial network is connectionless. Based on Pareto principle, 80% of the nodes represent universities with common level of knowledge and the remaining 20% represent universities with higher level of knowledge. The knowledge of node i is divided into λ knowledge dimensions ($\lambda = [1, 10]$) and each dimension is composed of explicit knowledge $X_{i,\lambda}$ and tacit knowledge $Y_{i,\lambda}$ according to the different attributes of knowledge.

It is assumed that nodes within the cluster innovation network have different levels of knowledge at the initial time; $X_{i,\lambda}$ and $Y_{i,\lambda}$ of the high-level universities take random values in the range of [0.8,1], and $X_{i,\lambda}$ and $Y_{i,\lambda}$ of the general-level universities take random values in the range of [0.4,0.8]. In the whole process of knowledge transfer, we separately take the average levels of explicit and tacit knowledge within 10 dimensions to simulate (as $\overline{X}_i$ and $\overline{Y}_i$).

As a special learning organization, universities' core competitiveness lies in tacit knowledge. Compared with explicit knowledge, universities have a relatively large proportion of tacit knowledge. Therefore, we assume that the overall knowledge level of node i is $\overline{Z}_i$: the weight of explicit knowledge is $m = 0.3$ and the weight of tacit knowledge is $(1 - m) = 0.7$.

$$\overline{X}_i = \frac{1}{10} \sum_{\lambda=1}^{10} X_{i,\lambda} \tag{1}$$

$$\overline{Y}_i = \frac{1}{10} \sum_{\lambda=1}^{10} Y_{i,\lambda} \tag{2}$$

$$\overline{Z}_i = m\overline{X}_i + (1 - m)\overline{Y}_i \tag{3}$$

The average knowledge level of the cluster innovation network as a whole is

$$Z_{AVG} = \sum_{i=1}^{N} \frac{\overline{Z}_i}{N} \tag{4}$$

3.2. The Simulation Model of Knowledge Transfer Process

3.2.1. Knowledge Externalization. In the process of externalization, the externalization efficiency of node i at time t is defined as $p_{i,t}$; $p_{i,t}$ of each university in the cluster innovation network at time t is equal in order to simplify the model. Specifically, tacit knowledge cannot be completely externalized because of the characteristics of universities' knowledge. This research assumes that the externalization efficiency $p_{i,t}$ is within the scope (0, 1). Meanwhile, with the continuous externalization of knowledge, $p_{i,t}$ will decrease to a stable limit value according to the following formula (5).

$$p_{i,t} = \eta e^{-\alpha t} \tag{5}$$

α is the externalization factor and different externalization factors (α) correspond to different externalization efficiencies ($p_{i,t}$). The larger α is, the lower externalization efficiency $p_{i,t}$ is. On account of universities with lower degree of tacit knowledge's externalization, the range of value for α is [0.7,1]. The value of adjustment coefficient is set to $\eta = 0.2$ through previous multiple tests and experiments.

In this stage, the levels of explicit and tacit knowledge and the externalization efficiency $p_{i,t}$ of node i at time t have a combined effect in explicit knowledge level of node i at the next moment $t + 1$. Therefore, the tacit knowledge externalization of node i is expressed as follows:

$$\overline{X}_{i,t+1} = \overline{X}_{i,t} + p_t \overline{Y}_{i,t} \tag{6}$$

3.2.2. Knowledge Sharing. The stage of knowledge sharing is mainly the exchange of explicit knowledge among subjects according to the principle of network connection based on knowledge potential differences. Some studies point out that only knowledge potential difference among innovation subjects in a reasonable range is an important driving force for knowledge transfer [34]. According to Huang Weiqiang [9], the cooperation and exchange between two knowledge subjects in the cluster should ensure that the comprehensive knowledge gap is within a suitable range. This paper uses Euclidean distance to express the comprehensive knowledge potential difference $D_{(i,j),t}$ between node i and j at time t.

$$D_{(i,j),t} = \sqrt{\left(\overline{X}_{j,t} - \overline{X}_{i,t}\right)^2 + \left(\overline{Y}_{j,t} - \overline{Y}_{i,t}\right)^2} \tag{7}$$

We propose δ_1 and δ, respectively, representing the lower limit and upper limit of the comprehensive knowledge potential difference. The knowledge potential differences of the cooperative universities must meet the upper and lower limits:

$$\delta_1 \leq D_{(i,j),t} \leq \delta \tag{8}$$

This study assumes that initial network is connectionless; that is, all nodes are independent of each other. Select a node i from nodes N and calculate the comprehensive knowledge potential difference between node i and the other node k in turn ($k \neq i$). Additionally, we suppose that the set of node k where $D_{(i,k),t}$ meet the upper and lower limits is A. At this time, the cooperative relationships between node i and nodes in the set A will be established, connecting node i and nodes in the set A to communicate and learn explicit knowledge. In addition, the set of nodes where explicit knowledge level of a

node in the set A is greater than node i is B. In the set B, we define the node j whose explicit knowledge level is the highest as $j_{\max}$, as shown in formula (9). The explicit knowledge level of node i can achieve ultimate level after exchanging explicit knowledge with node $j_{\max}$ follows the learning rule.

The explicit knowledge level of the node i at time $t + 1$ is associated with the learning ability of node i, the comprehensive knowledge potential difference between node i and node $j_{\max}$ at time t, and explicit knowledge levels of node i and $j_{\max}$ at time t. Therefore, on the premise of these assumptions, when subjects exchange and learn knowledge, the learning function of node i over time is defined as follows according to Huangweiqiang's research [9]:

$$\overline{X}_{j\max,t} = \max\left(\overline{X}_{j,t} \mid j \in B\right) \tag{9}$$

$$d_t = \sqrt{\left(\overline{X}_{j\max,t} - \overline{X}_{i,t}\right)^2 + \left(\overline{Y}_{j\max,t} - \overline{Y}_{i,t}\right)^2} \tag{10}$$

$$\overline{X}_{i,t+1} = \overline{X}_{i,t} - \theta\left(\overline{X}_{j\max,t} - \overline{X}_{i,t} - \frac{d_t}{2}\right)^2 + \frac{\theta\left(d_t\right)^2}{4} \tag{11}$$

In formula (11), d_t is the comprehensive knowledge potential difference between i and $j_{\max}$ at time t, $\overline{X}_{i,t}$ indicates the explicit knowledge level of node i at time t, and θ is the learning ability of node i.

This paper assumes that the knowledge level of knowledge receiver is less than that of knowledge sender. With the knowledge exchange, the explicit knowledge level of the receiver can be improved to a certain extent, but the maximum will not exceed the sender's level. At the same time, the level of explicit knowledge of the sender remains constant. When the nodes that meet the principle of network connection based on knowledge potential difference establish cooperative learning relationships, the knowledge levels of these nodes are improved, and the overall knowledge level of the cluster network is gradually improved. Meanwhile, the comprehensive knowledge potential between nodes changes, which will prompt them to break off the previous cooperative relationships and seek new partners to continue to learn and improve the level of knowledge. Thus, at this time, we disconnect all network relationships in order for it to restore the connectionless network. Then repeat the above operation until the comprehensive knowledge potential difference of all nodes in the cluster innovation network cannot meet the principle of connection, and knowledge learning and exchange between nodes stop. The knowledge level of each node in cluster innovation network gradually converges, and the average knowledge level of the entire network tends to be stable.

3.2.3. Knowledge Innovation.
The explicit knowledge learned has a subtle influence on the original explicit and tacit knowledge through knowledge exchange and sharing. Therefore, explicit and tacit knowledge will, respectively, innovate at this stage. This study supposes that the innovation ability of node i has a trend of diminishing marginal returns over time, and the explicit/tacit knowledge level of node i at time $t + 1$ is affected by explicit/tacit knowledge level and innovation

ability of node i at time t. According to Li Jinhua [35], the rule of explicit and tacit knowledge innovation of node i is as follows:

$$\overline{X}_{i,t+1} = \overline{X}_{i,t}\left(1 + e^{-\varphi(t-t')}\right) \tag{12}$$

$$\overline{Y}_{i,t+1} = \overline{Y}_{i,t}\left(1 + e^{-\varphi(t-t')}\right) \tag{13}$$

φ is the innovation factor of node i. Based on the knowledge characteristics of the university and previous multiple tests and simulation, this paper sets φ equal to 4. t' refers to the sum of observation periods of the first two stages of knowledge transfer (the knowledge externalization and knowledge sharing).

3.2.4. Knowledge Internalization.
After knowledge innovation, the university could continuously integrate and accumulate new explicit knowledge and apply it to the daily learning practice of teachers and students, thereby enhancing the core competence of insiders and internalizing explicit knowledge into noncoding tacit resource to improve the independent innovation ability of universities.

In this stage, the tacit and explicit knowledge level of node i and the internalization efficiency $q_{i,t}$ at time t can affect the tacit knowledge level of node i at time $t + 1$. Therefore, the rule of internalization is as follows:

$$\overline{Y}_{i,t+1} = \overline{Y}_{i,t} + q_{i,t}\overline{X}_{i,t} \tag{14}$$

$q_{i,t}$ refers to the knowledge internalization efficiency of node i at time t. In order to simplify the model, this paper assumes that the knowledge internalization efficiency of each node in the cluster innovation network has the same value at time t, and they take values within the range of $(0, 1)$. In addition, with the continuous internalization of knowledge, $q_{i,t}$ could decrease to a stable limit value according to the following formula (15).

$$q_{i,t} = \mu e^{-\beta(t-t'')} \tag{15}$$

β is an internalization factor that affects the efficiency of knowledge internalization. μ is a moderator. Based on the knowledge characteristics of universities and previous multiple tests and simulation, we assume $\mu = 0.03$ and the range of value for β is $[0.8, 1]$. t'' refers to the sum of observation periods of the first three stages of knowledge transfer (the stage of knowledge externalization, sharing, and innovation).

3.3. Network Topology Statistics.
Nowadays, many networks are becoming more and more complex. In order to expose the internal characteristics of these complex networks in detail, many scholars have proposed descriptive statistical indicators such as degree, degree distribution, clustering coefficient, and path length to reflect the network characteristics.

(1) Degree and Degree Distribution. The degree of node i is the number of other individuals connected with node i in

the network. Degree distribution refers to the distribution of degrees of all nodes in the entire network, recorded as $p(k)$.

(2) Average Clustering Coefficient. Clustering coefficient, a local feature of the network, reflects the clustering characteristics of the entire network. The degree of node i is k_i, that is, in the network the number of other nodes that have cooperative relationships with i is k_i. There are at most $k_i(k_i - 1)/2$ edges in these k_i nodes. The number of cooperation relationships between k_i nodes is e_i; that is, the number of edges that actually exists is e_i. At this time, the clustering coefficient of node i is C_i:

$$C_i = \frac{2e_i}{k_i(k_i - 1)} \tag{16}$$

The average clustering coefficient of the entire network is recorded as C, as shown in formula (17) (N is the number of network nodes):

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{17}$$

(3) Average Path Length. The minimum number of edges connecting arbitrary two nodes i and j in the entire network is the path length of these two nodes, recorded as $d_{i,j}$, and the average value of all $d_{i,j}$ in the network is the average path length, recorded as L:

$$L = \frac{1}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1} d_{i,j} \tag{18}$$

The average path length reflects the connectivity of the entire network. It is an important measurable index to describe the cooperation between the cross-cohesive subgroups. The more "cross-distance" connections, the more "shortcuts" of the network, and the network's average shortest path will be greatly reduced.

(4) Small World. According to some researches, many networks in the real world have small-world property; that is, the network has a high clustering coefficient and a short average path length. Davis et al. [36] compared the parameter index of the actual network with the parameter index of the random network with the same number of nodes and the number of links and proposed the small-world entropy, denoted as RSW.

$$RSW = \left(\frac{C_{actual}}{L_{actual}}\right) * \left(\frac{L_{random}}{C_{random}}\right) \tag{19}$$

If the average shortest path of the actual network and the random network are approximately equal and the clustering coefficient of the actual network is greater than the random network, the actual network has small-world property. In other words, when the small-world entropy is significantly greater than 1, we can judge that the actual network shows the small-world phenomenon.

4. Simulation Results

According to the above model, this paper quantitatively analyzes the mechanism of knowledge transfer in universities and how the cluster innovation network affects knowledge level using the numerical simulation method to further reveal the basic topology structure and dynamic evolution discipline of the cluster innovation network. The knowledge sharing stage emphasizes cooperative partner selection mechanisms and interactive learning mechanisms of different knowledge subjects, and this stage is also the key to the evolution of innovation network. At the same time, because of the paper's space limit, we only select the simulation process and results of knowledge externalization and knowledge sharing.

This study assumes that the innovation network is connectionless at initial moment, and there are 100 nodes in the network ($N = 100$). According to Baum J A C [37], the network's initial average knowledge level within knowledge sharing phase is different from initial average knowledge level of the literature [37], so this paper appropriately adjusts the range of the knowledge potential difference based on the research in the literature [37], and we take $\delta = 0.1, \delta_1 = 0.068$ through multiple tests.

The results of previous multiple tests and simulation show that the knowledge level of knowledge externalization and knowledge sharing stage could converge over a period of time. Therefore, we suppose that the total observation time of the first two stages of knowledge transfer is $T = 1300$ and the observation durations of the first two stages of knowledge transfer are $\Delta t_1 = 50, \Delta t_2 = 1250$, respectively. In order to eliminate single-shot errors as much as possible and reflect the evolutionary trend more scientifically, the simulation operation is repeated 20 times for each set of the parameters, and the final result is taken as the average of 20 simulation results.

4.1. Change Mechanisms of Network's Average Knowledge Level in Knowledge Externalization and Knowledge Sharing Stage

4.1.1. The Impact of Externalization Efficiency on the Network's Average Knowledge Level. Knowledge externalization is the initial stage of knowledge transfer within universities. According to the model, the externalization efficiency $p_{i,t}$ is affected by the externalization factor α. Therefore, the relationships between the externalization factor α and the average knowledge level of the cluster innovation network can reflect the impacts of the externalization efficiency $p_{i,t}$ on the network's average knowledge level. Meanwhile, due to the assumptions of the first stage and lower degree of tacit knowledge's externalization of universities, this paper supposes $\alpha = 0.7, 0.8, 0.9$ through previous multiple tests with 20 independent simulation operations, respectively, corresponding to three different externalization efficiencies $p_{i,t}$. Figure 2 shows that the larger the externalization factor α, the lower the knowledge externalization efficiency $p_{i,t}$. At this time, the average knowledge level of the innovation network tends to converge prematurely, and the convergence
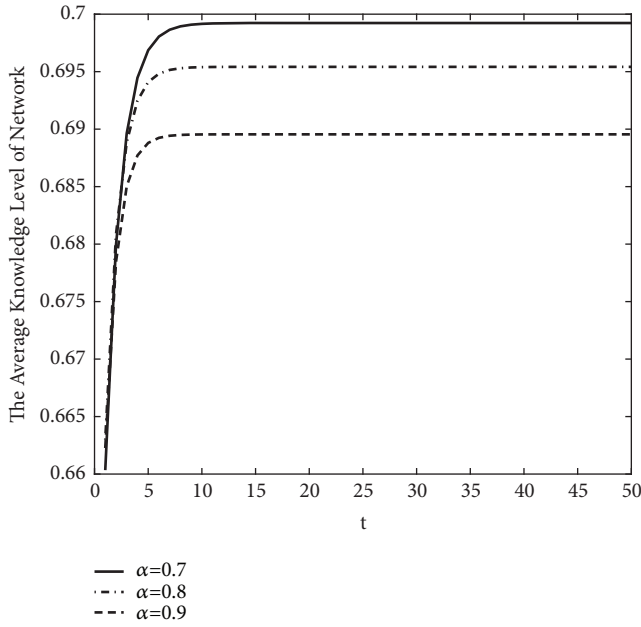
Figure 2: Network's average knowledge level evolution diagram under different α in knowledge externalization.

Figure 3: When $\alpha = 0.7$ network average knowledge level evolution diagram under different θ values.

value of knowledge level is lower. Overall, the externalization efficiency $p_{i,t}$ shows negative relation with the convergent speed of network average knowledge level, but $p_{i,t}$ is positively associated with the final convergence knowledge level.

4.1.2. The Impact of Learning Ability on the Average Knowledge Level of Network. With each university finishing externalization of tacit knowledge, the overall knowledge level of the innovation network can increase, and the second stage of knowledge sharing will begin after the first stage's knowledge level is stable and convergent.

In the process of knowledge sharing, the learning abilities of knowledge subjects are affected by many factors. Knowledge sharing is based on knowledge externalization in this paper; therefore, the average knowledge level of network in the final moment of knowledge externalization phase could affect the initial average knowledge level of knowledge sharing phase, thus impacting knowledge learning ability. In order to eliminate the contingency of research results, this study divides network's average knowledge level at initial moment of knowledge sharing stage into two sets of data for simulation according to externalization factor α; that is, we explore the impact of different learning abilities θ on network' average knowledge level when $\alpha = 0.7$ and $\alpha = 0.8$ to determine whether the two sets of results match.

When externalization factor $\alpha = 0.7$, the network's average knowledge level at the initial moment of knowledge sharing stage is around 0.7. We assume learning ability $\theta = 2, 5, 7, 8$ to analyze the evolving trend of the average knowledge level of innovation network under these four different learning abilities (Figure 3). It can be found that the knowledge levels of universities with different learning abilities show an increasing trend and converge to stable
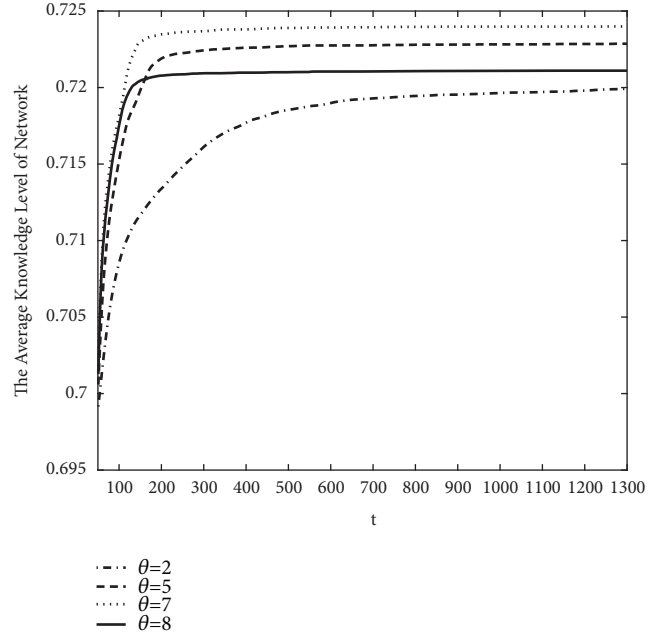
values, respectively, over time in Figure 3. When the learning ability is weak ($\theta = 2$), the network's average knowledge level converges the slowest. When the learning ability is strong ($\theta = 8$), the network's average knowledge level converges the fastest. The learning ability has a positive relationship with the convergence rate of the network's average knowledge level; that is, the stronger the learning ability, the faster the convergence rate. In particular, there is not any purely positive correlation between the learning ability and the final convergence value of network's average knowledge level. When $\theta < 8$, the larger the learning ability θ, the higher the final convergence level. But the convergence level when $\theta = 8$ is less than that when $\theta = 5$ and $\theta = 7$, which is around 0.721.

Figure 4 reveals the evolving trend of the average knowledge level of the cluster innovation network while $\alpha = 0.8$ and $\theta = 2, 5, 7, 8, 9$. The network's average knowledge level at the initial moment of knowledge sharing stage is around 0.696. We can find that the trend shown in Figure 4 is consistent with that in Figure 3. With time, the average knowledge levels of network with five different learning abilities show an increasing trend and converge to stable values, respectively, in Figure 4. The relationship between learning ability θ and average knowledge level of network is as follows: the learning ability is positively associated with the convergence speed of the network's average knowledge level. Additionally, the learning ability has a positive correlation with the final convergence value of network's average knowledge level when $\theta < 9$. However, the convergence level when $\theta = 9$ is less than that when $\theta = 8$; at this moment, there is no positive relationship between the learning ability and the final convergence level.

The two sets of simulation results above show the same phenomenon: the average knowledge levels of the cluster

TABLE 1: The network topology parameters under different θ values at $t = 100$ and $t = 800$.

topology structure parameters	t	$\theta = 2$		$\theta = 5$		$\theta = 8$	
		m	std	m	std	m	std
Average path length (L)	t=100	1.7891	0.0682	1.7636	0.1007	1.9871	0.2679
	t=800	2.0549	0.2608	2.3643	0.3661	2.4128	0.2851
Average clustering coefficient (C)	t=100	0.1725	0.0285	0.1739	0.0361	0.1802	0.0531
	t=800	0.1125	0.0504	0.0633	0.0255	0.0551	0.0220
Small world Q (RSW)	t=100	5.0311	0.2029	5.0315	0.1871	5.5323	0.3056
	t=800	2.3928	0.2842	0.6402	0.1781	0.3501	0.1760

Notes: (1) *m* (mean values)/*std* (standard deviation) are shown for 20 replications at each moment for the same parameters; (2) according to Davis G F [36], the average path length is calculated based on the maximal connected subgraphs of the innovation network. In the calculation formula of the small world quotient RSW, $L_{random} = \ln(N)/\ln(k)$, $C_{random} = k/N$.
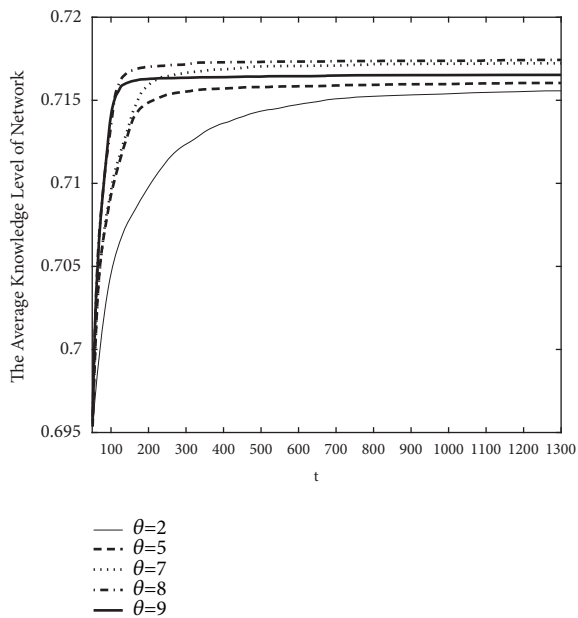
FIGURE 4: When $\alpha = 0.8$ network average knowledge level evolution diagram under different θ values.

innovation network under different learning abilities increase progressively with time and finally converge to steady values separately. It illustrates that knowledge sharing and learning enhance knowledge level and promote knowledge transfer performance. However, there is a gradual convergence of the knowledge levels of some universities in the cluster innovation network and other universities' knowledge levels have wider gap. At this time, the knowledge potential differences between each other are hard to meet the principle of network connection, so the cooperative relationships between universities cannot be established and the overall knowledge level of the innovation network reaches a saturated state.

In addition, Figures 3 and 4 show that the learning ability θ is positively correlated with the convergence speed of the innovation network's average knowledge level. On the contrary, the relationship between the learning ability and the final convergence value of the network's average knowledge level is not positive. Only when the learning ability θ is within a limited range, the larger the learning ability θ, the greater the final knowledge convergence level of the cluster innovation network. If the learning ability exceeds a certain value, the final convergence value of the network's average knowledge level drops off with the increase of the learning ability. This is because when knowledge subjects have stronger learning ability, they are satisfied with achieving the higher knowledge level quickly and lack motivation for learning new knowledge. On account of the emergence of "negative emotions," knowledge subjects will not effectively learn new knowledge even with the continuous evolution of the innovation network, so the average knowledge level of the innovation network reaches convergence.

4.2. Basic Topology of the Cluster Innovation Network. In the whole process of knowledge transfer, only did the knowledge sharing stage occur between different universities, and the other stages are internal activities of the university.

In particular, in the stage of knowledge sharing, the nodes establish learning cooperative relationships based on the principle of network connection-oriented knowledge potential difference over time. The knowledge levels of subjects and the cooperation relationships are constantly interacting, and cooperation relationships and network structures are constantly and intricately changing before the network's average knowledge level converges. In addition, according to the conclusions of Section 4.1, the knowledge learning ability θ is positively correlated with the final convergence value of network's average knowledge level only when θ is within a certain range. If θ exceeds this range, knowledge subjects should find another way to improve knowledge level. Therefore, in order to explore how the learning ability and knowledge level affect network topology and make the conclusions bring realistic significance, this paper only discusses the dynamic evolution law of network basic topology when learning ability θ is within the appropriate range in the stage of knowledge sharing.

Table 1 shows the network topology parameters under different learning abilities in the earlier period ($t = 100$) and the later period ($t = 800$) of knowledge sharing stage when externalization factor $\alpha = 0.8$ and learning ability $\theta = 2, 5, 8$. The data in Table 1 are the average of 20 simulation operations independently performed.
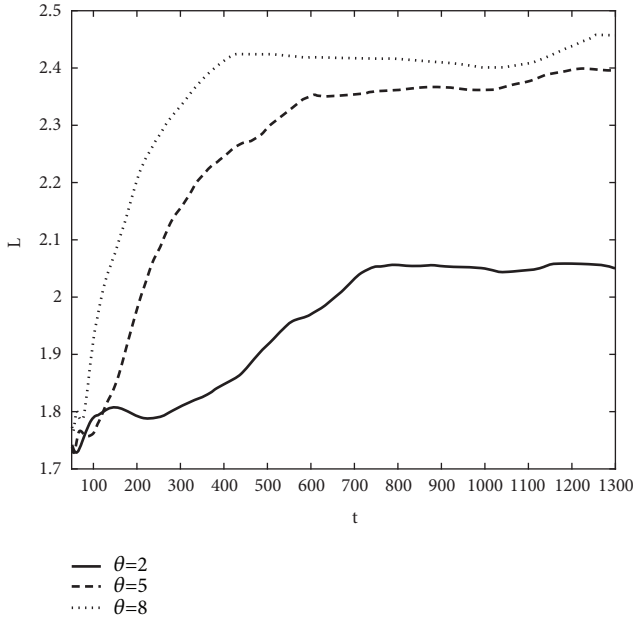
FIGURE 5: The evolution picture of the cluster innovation network average path length.
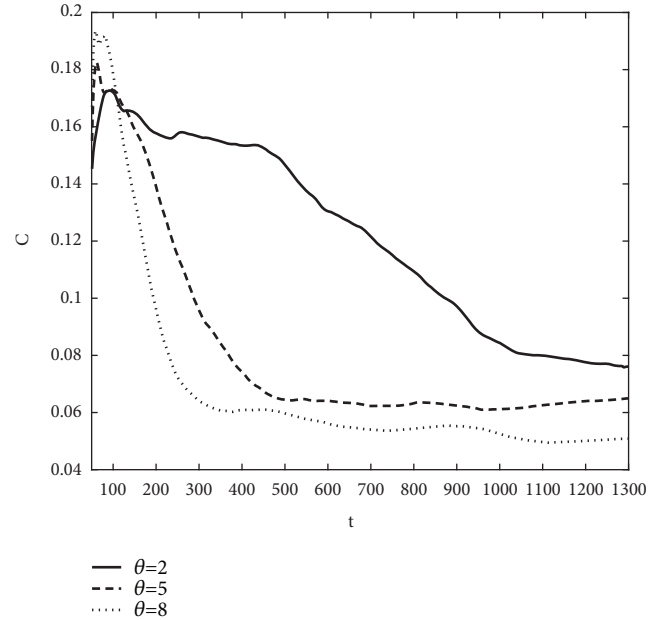
FIGURE 6: The evolution picture of the cluster innovation network average clustering coefficient.

Table 1 describes that the small world quotient (RSW) of each learning ability θ is greater than 1 at time $t = 100$, which indicates that the actual network shows the small-world phenomenon: $L_{actual} \approx L_{random}, C_{actual} >> C_{random}$. However, the small-world quotients under different θ at time $t = 800$ are less than those at time $t = 100$. We can deduce that the average clustering coefficient is small and the average path length is large in the later period of knowledge sharing stage. The innovation network gradually becomes sparse, and the network's small-world property weakens.

Table 1 is only a partial discussion of the network topology of the prenetwork and postnetwork in knowledge sharing stage. In order to explore network evolution rule more systematically and comprehensively, Figures 5, 6, and 7, respectively, depict the changes of the average path length, the average clustering coefficient, and small-world quotient of the innovation network when externalization factor $\alpha = 0.8$ and learning ability $\theta = 2, 5, 8$. To eliminate the influences of some uncertain factors and reveal the trend of each parameter more accurately, the data in Figures 5, 6, and 7 are the results of 20 moving averages of original data.

The commonalities of the network evolution trends under different θ values are as follows:

Figure 5 reveals that, with the evolution of the network in the process of knowledge sharing, the trends of the average path length of the innovation network under different learning abilities are approximately the same; the average path length increases to the first small peak at the beginning and decreases slightly soon after. Then the average path length continuously increases over time until it is stable at a higher level. Figures 6 and 7 show that, in the earlier period of knowledge sharing, the average clustering coefficients and the small-world quotients of different learning abilities
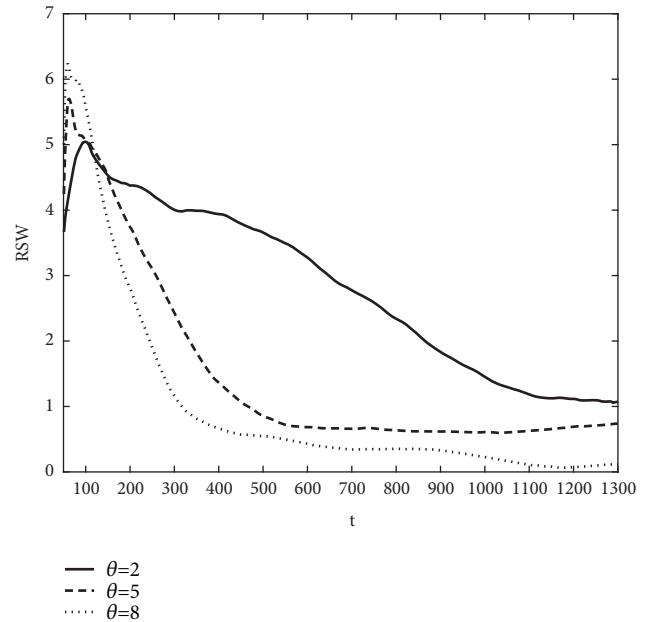
FIGURE 7: The evolution picture of the cluster innovation network Small-World quotient.

present consistently the inverse U evolution of the left-biased distribution, which increases firstly and then decreases. In the later period they have stabilized at lower levels, respectively.

We can combine Figures 5, 6, and 7 to explore the specific evolution of the university innovation network in the knowledge sharing phase from a global perspective:

(I) The First Period. In the early evolution, the average path length increases in a shorter period, but the path length at this

time was at a lower level relative to other moments. Meanwhile, the average clustering coefficient and the small-world quotient both show the trend of rapidly rising to the peak and then decreasing. In addition, the clustering coefficient's level at the early period is higher than other moments, and the small-world quotient is greater than 1. These illustrate that, in the early stage, there are a large number of universities with cooperative relationships in the cluster innovation network, and the network has a relatively high degree of clustering. Although a minority of key "remote-cooperations" turn into "adjacent-cooperations" gradually, which makes the average path length of the network increase slightly, the overall level of average path length is still relatively short. Therefore, the innovation network presents the particularly obvious small-world property during this period.

(II) The Second Period. The average path length begins to show a small amplitude short-term decline after the initial small-range rise. Meanwhile, the average clustering coefficient and the small-world quotient are still in the trend of decline. During this period, the network has a relatively high degree of clustering and a short average path length, and the small-world quotient is still significantly larger than 1. From the perspective of the shorter average path length and the higher average clustering coefficient, in this period, the number of "adjacent-cooperations" in the network reduces slightly, and the number of "remote-cooperations" has a small increase, which could lead to a decrease in the average path length and an increase in the degree of clustering. The network shows the significant small-world phenomenon.

(III) The Third Period. At this period, the average path length of the network begins to rise, but the increase velocity in the later stage is getting smaller and smaller. On the contrary, the average clustering coefficient and the small-world quotient decrease, and the rate of decline in the later stage is slower and slower. At the same time, small-world quotient gradually becomes less than 1 in the process of decline. These show that the number of "remote-cooperations" and "adjacent-cooperations" decreases in varying degrees. In addition, the degree of network clustering reduces, and the compactness of cooperation relationships between universities is relatively weak. Overall, the long average path length and the low average clustering coefficient indicate that the efficiency of knowledge dissemination in the cluster network is greatly reduced, and the small-world network is gradually disrupted.

In summary, in the knowledge sharing stage, with the continuous learning of universities, the innovation network is constantly evolving and shows a significant small-world phenomenon in the process of evolution.

4.3. Collaborative Evolution of Knowledge Level and Innovation Network Structure. Universities, important subjects of innovation network, decide whether to conduct learning cooperation relationships between each other according to the knowledge potential difference principle. In the knowledge sharing stage, the knowledge levels of subjects can affect the comprehensive knowledge potential differences,

further influencing the cooperative decision-making processes and the evolution of the innovation network structure. Meanwhile, cooperation relationships and network structure change, which in turn affect the knowledge growth performance and the average knowledge level of the network.

It can be found from Figures 3, 5, 6, and 7 that the rising period of the average knowledge level of the innovation network is the period that small-world property of the network is significant, and the period when the average knowledge level of the network tends to be stable is the period when the small-world network gradually begins to collapse.

When the network with a relatively high clustering coefficient and small characteristic path length shows the conspicuous small-world phenomenon, the cluster innovation network cooperation is highly clustered relatively and the network distance is short. At this moment, knowledge and information transmission have high efficiency to achieve high performance of knowledge sharing. It is helpful to knowledge learning, which can further improve the average knowledge level of the network. When the small-world property begins to be inconspicuous or even gradually disintegrating, the average path length increases and the average clustering coefficient decreases. At this time, the number of "adjacent-cooperations" is more than the number of "remote-cooperations." The network is gradually sparse. It is not conducive to knowledge learning and sharing, and the information transmission is less efficient, which could result in the average knowledge level of the network not ascend.

Conversely, the level of knowledge also affects the network topology. In the early period of knowledge sharing, the explicit knowledge levels of universities in the network are relatively low, so they have a strong desire to improve their knowledge level. At the same time, because the comprehensive knowledge potential differences of most universities meet the conditions of cooperation, they have a strong connectivity base to establish cooperative relationships. These connections make the innovation network with the small-world property: a short average path length and a high average clustering coefficient.

After a period of knowledge learning and sharing, knowledge levels of many universities are gradually improved and converge to a stable condition. At this time, the knowledge potential differences between each other are too big or too small to meet the conditions for cooperation. The dissemination of knowledge is hindered, which can aggravate the collapse of the small-world network.

Overall, there is a coevolutionary relationship between the knowledge level of university and the innovation network structure in the stage of knowledge sharing.

5. Conclusion

In order to promote the construction of "Double First-Rate," strengthening the innovation cooperation of universities is an important way to improve the abilities of independent innovation. This requires constant knowledge interaction and transfer between universities in the cluster innovation network, and these processes could promote the continuous

evolution of the network. This paper combines the attribute dimension of knowledge (explicit and tacit knowledge), the principle of network connection based on knowledge potential difference and knowledge transfer process to establish a four-stage knowledge transfer model within universities cluster innovation network, and depicts the conversion influence mechanism of explicit and tacit knowledge and the internal mechanism of knowledge transfer quantitatively using the system simulation method to further explore the basic topology and dynamic evolution rules of the cluster innovation network in the knowledge sharing phase. The study draws the following conclusions:

(1) In the stage of knowledge sharing, each university establishes the cooperative relationship based on the principle of network connection-oriented knowledge potential difference for knowledge learning. The network structure evolves constantly with the change of cooperative relationships. In addition, the innovation network shows a significant small-world property in the dynamic evolution process. In the early period of knowledge sharing, the network has a short average path length and a relatively high degree of clustering; meanwhile, the small-world quotient is significantly larger than 1. Therefore, the innovation network shows the obvious small-world phenomenon. After a period of interaction and learning, the cooperation relationships of knowledge subjects become very sparse, and the number of "remote-cooperations" is gradually reduced. At this moment, the small-world quotient of the network with a long average path length and a low average clustering coefficient is less than 1. The small-world network begins to gradually collapse.

(2) There is a coevolutionary relationship between the knowledge level of university and the innovation network: in the knowledge sharing stage, many universities improve the levels of explicit knowledge through learning and cooperation and then improve their overall knowledge levels. At this time, the knowledge gaps between some universities will be widened or narrowed, which affects the choices of partners and the evolution of the innovation network. Meanwhile, the network evolution can change the path length and clustering coefficient of the network. The short average path length and high average clustering coefficient appeared in the evolution enhancing the compactness of network cooperation, which greatly improves the efficiency of knowledge transfer and learning, and further affecting the average knowledge level of the network.

(3) In the knowledge externalization stage, there are certain correlations between the externalization efficiency and the average knowledge level of the innovation network: The network's average knowledge level increases at first and then converges to a stable value in the process of externalization of tacit knowledge. Meanwhile, the knowledge externalization efficiency is negatively correlated with the convergence speed of the network's average knowledge level and positively associated with the final convergence value. Therefore, universities should improve the efficiency of knowledge externalization through reducing the externalization factor appropriately, thus improving the average knowledge level of the innovation network. For example, in the stage of knowledge externalization, under the premise of not infringing on intellectual property, university researchers can perfect the process of knowledge coding and establish corresponding knowledge databases actively to distill the experience and knowledge accumulated in the ordinary practice process into a clear knowledge mapping structure and so on for the exchange and sharing of knowledge in the next stage.

(4) In the knowledge sharing stage, the average knowledge levels of the network under different learning abilities show a similar trend, increase, and after a certain period of time converge to a stable value. In addition, the learning ability has a positive relationship with the convergence speed of the network's average knowledge level. Specifically, only when the learning ability is within a certain range, it is positively correlated with the final convergence value of the network's average knowledge level. If this range exceeded, this positive relationship does not exist.

Therefore, while universities improve their learning abilities actively, they should pay attention to the fact that when the learning ability is raised to a certain level, the knowledge agents lack motivation to learn new knowledge for what they think that their levels are high enough, because they reach the higher knowledge level relatively quickly. The average knowledge level of the cluster innovation network reaches a saturated state. At this moment, universities should seek other new partners outside the cluster actively to continue establishing cooperative relationships.

Data Availability

The simulation data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] S. Tallman, M. Jenkins, N. Henry, and S. Pinch, "Knowledge, clusters, and competitive advantage," *Academy of Management Review (AMR)*, vol. 29, no. 2, pp. 258–271, 2004.

[2] Y.-L. Lai, M.-S. Hsu, F.-J. Lin, Y.-M. Chen, and Y.-H. Lin, "The effects of industry cluster knowledge management on innovation performance," *Journal of Business Research*, vol. 67, no. 5, pp. 734–739, 2014.

[3] B. T. Asheim and A. Isaksen, "Regional innovation systems: the integration of local 'sticky' and global 'ubiquitous' knowledge," *The Journal of Technology Transfer*, vol. 27, no. 1, pp. 77–86, 2002.

[4] X. Xie, L. Fang, and S. Zeng, "Collaborative innovation network and knowledge transfer performance: A fsQCA approach," *Journal of Business Research*, vol. 69, no. 11, pp. 5210–5215, 2016.

[5] F. Wei, D. Sheng, and W. Lili, "Evolutionary model and simulation research of collaborative innovation network: a case study of artificial intelligence industry," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 4371528, 13 pages, 2018.

[6] L. I. Min, Y. L. Song, and L. I. Zi-Ting, "Knowledge transfer networks in cluster based on san—the case of foshan lighting cluster," *Systems Engineering*, vol. 35, no. 1, pp. 46–57, 2017.

[7] W. Fang, L. Tang, P. Cheng, and N. Ahmad, "Evolution decision, drivers and green innovation performance for collaborative innovation center of ecological building materials and environmental protection equipment in jiangsu province of china," *International Journal of Environmental Research and Public Health*, vol. 15, no. 11, p. 2365, 2018.

[8] Z. He, L. Rayman-Bacchus, and Y. Wu, "Self-organization of industrial clustering in a transition economy: A proposed framework and case study evidence from China," *Research Policy*, vol. 40, no. 9, pp. 1280–1294, 2011.

[9] H. Weiqiang, Z. Xintian, and Y. Shuang, "Research on the evolution of enterprise cluster innovation network based on dynamic knowledge complementation," *Scientific Research*, vol. 29, no. 10, pp. 1557–1567, 2011.

[10] M. Lin and N. Li, "Scale-free network provides an optimal pattern for knowledge transfer," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 3, pp. 473–480, 2010.

[11] T. G. Bunnell and N. M. Coe, "Spaces and scales of innovation," *Progress in Human Geography*, vol. 25, no. 4, pp. 569–589, 2001.

[12] X. M. Xie, L. L. Zuo, S. X. Zeng, and V. W. Y. Tam, "The impacts of network structures and network form on corporate innovative performance: evidence from high-tech sectors," *Asian Journal of Technology Innovation*, vol. 22, no. 2, pp. 185–203, 2014.

[13] I. M. Taplin, "Network structure and knowledge transfer in cluster evolution: The transformation of the Napa Valley wine region," *International Journal of Organizational Analysis*, vol. 19, no. 2, pp. 127–145, 2011.

[14] M. Fritsch and M. Kauffeld-Monz, "The impact of network structure on knowledge transfer: An application of social network analysis in the context of regional innovation networks," *Annals of Regional Science*, vol. 44, no. 1, pp. 21–38, 2009.

[15] H. Kim and Y. Park, "Structural effects of R&D collaboration network on knowledge diffusion performance," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8986–8992, 2009.

[16] B. He and G. Song, "Model analysis of tacit knowledge transfer based on cluster network structure features," *Advances in Information Sciences and Service Sciences*, vol. 4, no. 9, pp. 247–253, 2012.

[17] X. Wang, "Forming mechanisms and structures of a knowledge transfer network: theoretical and simulation research," *Journal of Knowledge Management*, vol. 17, no. 2, pp. 278–289, 2013.

[18] M. A. Xuejun, W. U. Jie, and Y. Zhang, "Research on industry alliance knowledge transfer network modeling and simulation based on complex networks," *Management Science and Engineering*, vol. 7, no. 3, pp. 13–21, 2013.

[19] D. J. Teece, "Technology transfer by multinational firms: the resource cost of transferring technological know-how," *The Economic Journal*, vol. 87, no. 346, p. 242, 1977.

[20] I. Nonaka and H. Takeuchi, *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, Oxford, UK, 1995.

[21] J. Chen and H. U. Hanhui, "Research on knowledge transfer in industrial cluster network—a case from daminglu automobile sale and service cluster," *Science of Science and Management of S and T*, vol. 2, pp. 100–104, 2010.

[22] C. E. Prescott, "Innovation in the multinational firm with globally dispersed R&D: technological knowledge utilization and accumulation," *The Journal of High Technology Management Research*, vol. 10, no. 2, pp. 203–221, 1999.

[23] M. Sakakibara, "Knowledge sharing in cooperative research and development," *Managerial and Decision Economics*, vol. 24, no. 2-3, pp. 117–132, 2003.

[24] J. Ulhi, H. Neergaard, and T. Bjerregaard, "Beyond unidirectional knowledge transfer an empirical study of trust-based university-industry research and technology collaboration," *International Journal of Entrepreneurship and Innovation*, vol. 1304, pp. 287–299, 2012.

[25] C. Ryu, Y. J. Kim, A. Chaudhury, and H. R. Rao, "Knowledge acquisition via three learning processes in enterprise information portals: Learning-by-investment, learning-by-doing, and learning-from-others," *MIS Quarterly: Management Information Systems*, vol. 29, no. 2, pp. 245–278, 2005.

[26] M. C. J. Caniëls and B. Verspagen, "Barriers to knowledge spillovers and regional convergence in an evolutionary model," *Journal of Evolutionary Economics*, vol. 11, no. 3, pp. 307–329, 2001.

[27] Z. Gupeng, "The dynamic evolution of the small world innovation network and its effects," *Journal of Management Science*, vol. 18, no. 6, pp. 15–29, 2015.

[28] F. Karnani, "The university's unknown knowledge: Tacit knowledge, technology transfer and university spin-offs findings from an empirical study based on the theory of knowledge," *The Journal of Technology Transfer*, vol. 38, no. 3, pp. 235–250, 2013.

[29] M. Marra, "The contribution of evaluation to socialization and externalization of tacit knowledge: the case of the world bank," *Evaluation*, vol. 10, no. 3, pp. 263–283, 2004.

[30] S. S. Rao and A. Nayak, "EO model for tacit knowledge externalization in socio-technical enterprises," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 12, pp. 99–124, 2017.

[31] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[32] T. van der Valk, M. M. H. Chappin, and G. W. Gijsbers, "Evaluating innovation networks in emerging technologies," *Technological Forecasting & Social Change*, vol. 78, no. 1, pp. 25–39, 2011.

[33] F. Hermans, D. Van Apeldoorn, M. Stuiver, and K. Kok, "Niches and networks: Explaining network evolution through niche formation processes," *Research Policy*, vol. 42, no. 3, pp. 613–623, 2013.

[34] M. Han, "Research on simulation for the mechanism of knowledge potential difference between enterprises in the industrial cluster on knowledge transfer," *Applied Mechanics and Materials*, vol. 246-247, pp. 317–321, 2013.

[35] L. Jinhua, "The impact of knowledge flow on innovation network structure—based on complex network theory," *Science and Technology Progress and Policy*, vol. 24, no. 11, pp. 91–94, 2007.

[36] G. F. Davis, M. Yoo, and W. E. Baker, "The small world of the american corporate elite, 1982-2001," *Strategic Organization*, vol. 1, no. 3, pp. 301–326, 2003.

[37] J. A. C. Baum, R. Cowan, and N. Jonard, "Network-independent partner selection and the evolution of innovation networks," *Management Science*, vol. 56, no. 11, pp. 2094–2110, 2010.

Research Article

Predicting Missing Links Based on a New Triangle Structure

Shenshen Bai,[1,2] Longjie Li [ID],[1] Jianjun Cheng,[1] Shijin Xu,[1] and Xiaoyun Chen [ID][1]

[1]*School of Information Science & Engineering, Lanzhou University, Lanzhou 730000, China*
[2]*Department of Electronic and Information Engineering, Lanzhou Vocational Technical College, Lanzhou 730070, China*

Correspondence should be addressed to Longjie Li; ljli@lzu.edu.cn and Xiaoyun Chen; chenxy@lzu.edu.cn

With the rapid growth of various complex networks, link prediction has become increasingly important because it can discover the missing information and predict future interactions between nodes in a network. Recently, the CAR and CCLP indexes have been presented for link prediction by means of different triangle structure information. However, both indexes may lose the contributions of some shared neighbors. We propose in this work a new index to make up the weakness and then improve the accuracy of link prediction. The proposed index focuses on a new triangle structure, i.e., the triangle formed by one seed node, one common neighbor, and one other node. It emphasizes the importance of these triangles but does not ignore the contribution of any common neighbor. In addition, the proposed index adopts the theory of resource allocation by penalizing large-degree neighbors. The results of comparison with CN, AA, RA, ADP, CAR, CAA, CRA, and CCLP on 12 real-world networks show that the proposed index outperforms the compared methods in terms of AUC and ranking score.

1. Introduction

As a fundamental research hotspot in complex network analysis, link prediction has a wide range of applications in both theory and reality, such as analysis of network evolution [1, 2], recommendation system [3], and checking potential interactions between proteins in biological networks [4, 5]. The basic task of link prediction is to estimate the missing or latent existent links between unconnected nodes in a network [6, 7]. To date, a host of algorithms and models have been proposed for link prediction [6, 8, 9]. Reference [8] groups them into two ways: *similarity-based approaches* and *learning-based approaches*. A similarity-based approach computes similarity scores between unconnected nodes based on the known information. Then, a ranked list of node pairs in descending order according to their similarity scores is obtained and the node pairs at the top are thought most likely to have links. A learning-based approach formalizes the link prediction problem into a binary classification task [10] and uses machine learning methods to solve the problem. The key job in a learning-based approach is to construct the feature vectors of node pairs. In general, learning-based approaches are more complicated than similarity-based ones.

The hypothesis behind similarity-based approaches is *the more similar that two nodes are, the more likely that a link exists between them* [8]. This idea is simple and intuitive. Thus, the study of this kind of approaches has become the mainstream [6, 9]. The Common Neighbors (CN) index [11], as its name suggests, simply counts the number of common neighbors between two nodes. The Adamic-Adar (AA) [12] and Resource Allocation (RA) [13] indexes are two variants of the CN index; they penalize the contributions of large-degree common neighbors. These indexes are called local methods because they only use local structure information. Besides, some global and quasilocal methods have also been proposed by researchers, such as Katz [14], SimRank [15], Random Walks with Restart [16], Local Path [17], FriendLink [18], and Local Random Walk [19].

With the increasing growth of sizes of complex networks, local methods are still good candidates because they are more efficient in terms of running time than global and quasilocal methods. Therefore, we focus in this study on local methods. Recently, Cannistraci *et al.* proposed the CAR index [20], which suggests that links between the common neighbors, i.e., *local-community-links* (LCLs), are more valuable than common neighbors in link prediction. In CAR index, a *local*

(a) Example network

(b) Triangles used in CAR index

(d) Triangles used in TRA index

(c) Triangles used in CCLP index

$CN(a,b) = 4$ $CCLP(a,b) = \frac{4}{3}$

$RA(a,b) = \frac{4}{3}$ $TRA(a,b) = \frac{49}{24}$

$CAR(a,b) = 4$

(d) Similarity of (a, b)

● seed node

◐ non-common neighbor

◐ common neighbor

◐ neighbor of common neighbor
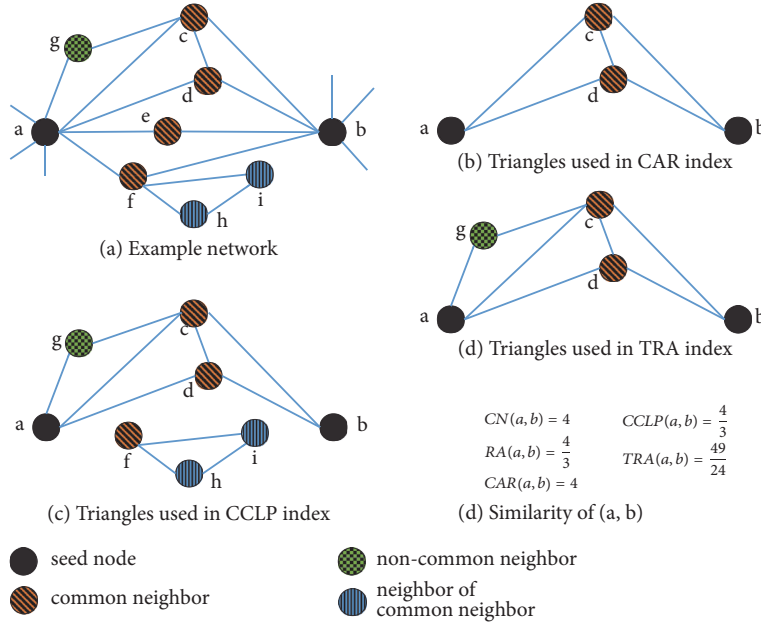
FIGURE 1: Triangles used in similarity indexes.

community is a triangle passing through two common neighbors and one seed node. In the example network shown in Figure 1(a), there is one LCL between the common neighbors of seed nodes a and b (see Figure 1(b)). Thus, CAR index assigns a similarity score of four to nodes a and b. However, if we remove the link between c and d, CAR will assign a zero similarity score to a and b, even though they have four common neighbors. In addition, the idea of LCL is also plugged into AA, RA, and Jaccard indexes [20]. Later, Wu *et al.* proposed the CCLP index based on the clustering coefficients of common neighbors. This index considers all triangles passing through a common neighbor. For the example network in Figure 1(a), there are triangles passing through nodes c, d, and f, respectively (see Figure 1(c)). Thus, CCLP index accumulates the clustering coefficients of nodes c, d, and f when calculating the similarity between a and b, but utterly neglects the contribution of node e. In real-world networks, it is possible that there are no triangles passing through some or even all shared neighbors of one node pair. Thus, CAR and CCLP indexes may assign a very low or even zero similarity score to the node pair, even if it has many common neighbors.

In this paper, we defines a new type of triangle structure, called TRA-triangle, which is formed by one seed node, one common neighbor, and one other node (see Figure 1(d)). Based on the TRA-triangle, a new similarity index, namely, TRA index, is proposed for link prediction. This index suggests that the common neighbors that can form TRA-triangles with a seed node are more important than others. In addition, the proposed index also penalizes the large-degree neighbors, as done in RA index [13]. Although all the TRA, CAR-based, and CCLP indexes are based on triangle structures, the intuitions behind them are different. The CAR-based indexes believe that LCLs are more valuable than

common neighbors. The CCLP index is inspired by CAR index but employs all triangles passing through common neighbors, while the TRA index, which only uses the TRA-triangles, strikes a balance between CAR and CCLP. Furthermore, as aforementioned, CAR-based and CCLP indexes lose the contribution of those common neighbors with no triangles passing through them, whereas TRA index counts the contribution of all kinds of common neighbors. Therefore, TRA index can achieve better prediction accuracy than CAR-based indexes and CCLP index. The accuracy of TRA index is evaluated on 12 real-world networks from various fields. The experimental results show that our index is far superior to CAR-based indexes and CCLP index. Take the network of HEP as an example, which is a very sparse network, the improvements made by TRA on CAR and CCLP, under the metric of AUC, are up to 26.9% and 4.2%, respectively.

The rest of the paper is structured as follows. In Section 2, we give the description of the link prediction problem and the evaluation metrics, list the compared methods and networks, and depict the Wilcoxon signed-ranks test. Section 3 introduces the proposed method. In Section 4, the experimental results and performance analysis of the proposed method are presented. Finally, Section 5 concludes this work.

2. Preliminaries

2.1. Problem Description and Metric. Given an undirected and unweighted network $G(V, E)$, in which V and E are the node set and link set, respectively, in this study, multilinks and self-loops are not allowed. Let $N = |V|$ be the number of nodes in the network, and let U be the universal possible link set, which contains $N(N-1)/2$ possible links. Then, the set of nonobserved links or nonexisting links is $U-E$. Suppose there are some missing links in $U - E$, the task of link prediction

is to find those links. A similarity-based approach assigns a similarity score to each node pair in $U - E$ and assumes that the higher score a node pair has, the more likely there is a link between them.

To test the performance of a similarity index, we randomly divide the link set E into two parts: training set E_{tr} and testing set E_{ts}, such that $E = E_{tr} \cup E_{ts}$ and $E_{tr} \cap E_{ts} = \emptyset$. E_{tr} is supposed to be the observed information, and E_{ts} is used for testing. Two parameter-free metrics are employed to quantify the accuracy of link prediction algorithms: AUC [6] and ranking score [21, 22]. In this situation, the AUC score can be interpreted as the probability that a randomly selected missing link (i.e., a link in E_{ts}) is given a higher score than a randomly selected nonexistent link (i.e., a link in $U - E$). When implementing, if we perform n independent comparisons, there are n_1 times that the missing link has higher score and n_2 times that they have the same score. The AUC value is then computed as

$$AUC = \frac{n_1 + 0.5n_2}{n}. \tag{1}$$

Ranking score (RS) takes the ranks of links in testing set after sorting in descend order according to their similarity scores into consideration. Let $H = U - E_{tr}$ be the set of nonobserved links. Let e_i be a missing link in E_{ts} and r_i be its rank. The ranking score of e_i is defined as $RS(e_i) = r_i/|H|$, and the ranking score of the link prediction result is as follows:

$$RS = \frac{1}{|E_{ts}|} \sum_{e_i \in E_{ts}} RS(e_i) = \frac{1}{|E_{ts}|} \sum_{e_i \in E_{ts}} \frac{r_i}{|H|}. \tag{2}$$

Note that the AUC value is the higher the better, whereas the ranking score is the smaller the better.

2.2. Local Similarity Indexes. As yet, many similarity indexes have been proposed for link prediction [6, 8, 9]. Here, we list some local similarity indexes that will be used in our experiments for the purpose of comparison.

(1) Common Neighbor (CN) index [11] defines the similarity between x and y as the number of their common neighbors, which is

$$CN(x, y) = |\Gamma(x) \cap \Gamma(y)|, \tag{3}$$

where $\Gamma(x)$ denotes the set of neighbors of node x.

(2) Adamic-Adar (AA) index [12] is a variant of CN index, which believes that small-degree neighbors have more contributions than large-degree neighbors when computing similarity. Its definition is as follows:

$$AA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log k_z}, \tag{4}$$

where k_z is the degree of node z.

(3) Resource Allocation (RA) index [13] defines the similarity between x and y as the amount of resource that y received from x through their common neighbors, which is

$$RA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z}. \tag{5}$$

(4) Adaptive Degree Penalization (ADP) index [23] penalizes a common neighbor according to its degree and the average clustering coefficient of the network. Therefore, it can automatically adapt to the network. The definition of ADP index is as follows:

$$ADP(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} k_z^{-\beta C}, \tag{6}$$

where β is a constant and C is the average clustering coefficient of the network. We set $\beta = 2.5$, as suggested by the authors.

(5) CAR index [20] suggests that two seed nodes are more likely to link together if there are links between their common neighbors, which is defined as

$$CAR(x, y) = CN(x, y) \cdot \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{L(z)}{2}, \tag{7}$$

where $L(z)$ is the number of links between z and other common neighbors of x and y.

(6) CAA and CRA indexes [20] are generated by plugging the idea of CAR index into the AA and RA indexes, respectively, which are defined as

$$CAA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{L(z)}{\log k_z}, \tag{8}$$

$$CRA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{L(z)}{k_z}. \tag{9}$$

(7) CCLP index [24] computes the similarity between x and y by employing clustering coefficient of common neighbors, which is

$$CCLP(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} CC_z, \tag{10}$$

where CC_z denotes the clustering coefficient of node z, which is

$$CC_z = \frac{2t_z}{k_z(k_z - 1)}, \tag{11}$$

in which t_z is the number of triangles passing through node z.

2.3. Networks. In this study, we use 12 real-world networks drawn from various fields to evaluate the effectiveness of link prediction methods.

(1) Advogato (ADV): a social network whose users are mainly free and open source software developers [25].

(2) C.elegans (CE): the neural network of a Caenorhabditis elegans worm [26].

(3) Dolphin: a social network of 62 dolphins in a community living off Doubtful Sound, New Zealand [27].

(4) Email: a network of email interchanges between members of a university [28].

(5) Foodweb (FW): a food web in Florida Bay during the rainy season [29].

(6) Hamster: a friendship network between users on hamsterster.com [30].

(7) HEP: the coauthorships network of scientists who posted preprints on the high-energy theory archive from 1995 to 1999 [31].

(8) Karate: the social network of a karate club at a US university [32].

(9) Political blogs (PB): a network of blogs about US politics [33].

(10) USAir: a network of the US air transportation system [6].

(11) Word: an adjacency network of common adjectives and noun in the novel "David Copperfield" by Charles Dickens [34].

(12) Yeast: the protein-protein interaction network of budding yeast [35].

In this work, all the aforementioned networks are treated as undirected and unweighted networks, and only the giant component of each network is used. Table 1 lists the basic statistics of the giant components of these networks.

Given network $G(V, E)$, suppose x, y be two seed nodes. (x, y) is called *a seed node pair with common neighbors* if they have at least one common neighbor. P_Λ denotes the set of seed node pairs with common neighbors, formally

$$P_\Lambda = \{(x, y) \mid (x, y) \notin E \wedge \Gamma(x) \cap \Gamma(y) \neq \emptyset\}. \quad (12)$$

Let x, y be two seed nodes, and z is one of their common neighbors. If $CC_z = 0$, we call z is a *zero-triangle-neighbor*; otherwise, z is a *triangle-neighbor*. If $L(z) \neq 0$, z is called a *CAR-triangle-neighbor* and if $\triangle(x, y; z) \neq 0$ (see (18)), z is called a *TRA-triangle-neighbor*. Let $S_\triangle$ be the set of triangle-neighbors, and S_{CAR}, S_{TRA} denote the sets of CAR- and TRA-triangle-neighbors, respectively. Clearly, $S_{CAR} \subseteq S_{TRA} \subseteq S_\triangle$. Let $P_\exists(\overline{S}_\triangle)$ and $P_\forall(\overline{S}_\triangle)$ be two subsets of P_Λ. For any pair in $P_\exists(\overline{S}_\triangle)$, at least one of their shared neighbors is not a triangle-neighbor, and for any pair in $P_\forall(\overline{S}_\triangle)$, all of their shared neighbors are not triangle-neighbors. More explicitly,

$$
\begin{aligned}
&P_\exists\left(\overline{S}_\triangle\right) \\
&= \{(x, y) \in P_\Lambda \mid \exists z \in \Gamma(x) \cap \Gamma(y) \wedge z \notin S_\triangle\}, \\
&P_\forall\left(\overline{S}_\triangle\right) \\
&= \{(x, y) \in P_\Lambda \mid \forall z \in \Gamma(x) \cap \Gamma(y) \wedge z \notin S_\triangle\}.
\end{aligned} \quad (13)
$$

Similarly, we define $P_\exists(\overline{S}_{TRA})$, $P_\forall(\overline{S}_{TRA})$, $P_\exists(\overline{S}_{CAR})$, and $P_\forall(\overline{S}_{CAR})$, which are

$$
\begin{aligned}
&P_\exists\left(\overline{S}_{TRA}\right) \\
&= \{(x, y) \in P_\Lambda \mid \exists z \in \Gamma(x) \cap \Gamma(y) \wedge z \notin S_{TRA}\}, \\
&P_\forall\left(\overline{S}_{TRA}\right) \\
&= \{(x, y) \in P_\Lambda \mid \forall z \in \Gamma(x) \cap \Gamma(y) \wedge z \notin S_{TRA}\}, \\
&P_\exists\left(\overline{S}_{CAR}\right) \\
&= \{(x, y) \in P_\Lambda \mid \exists z \in \Gamma(x) \cap \Gamma(y) \wedge z \notin S_{CAR}\}, \\
&P_\forall\left(\overline{S}_{CAR}\right) \\
&= \{(x, y) \in P_\Lambda \mid \forall z \in \Gamma(x) \cap \Gamma(y) \wedge z \notin S_{CAR}\}.
\end{aligned} \quad (14)
$$

Correspondingly, the ratios of those subsets to P_Λ are, respectively, defined as

$$
\begin{aligned}
R_\exists\left(\overline{S}_\triangle\right) &= \frac{\left|P_\exists\left(\overline{S}_\triangle\right)\right|}{|P_\Lambda|}, \\
R_\forall\left(\overline{S}_\triangle\right) &= \frac{\left|P_\forall\left(\overline{S}_\triangle\right)\right|}{|P_\Lambda|}, \\
R_\exists\left(\overline{S}_{TRA}\right) &= \frac{\left|P_\exists\left(\overline{S}_{TRA}\right)\right|}{|P_\Lambda|}, \\
R_\forall\left(\overline{S}_{TRA}\right) &= \frac{\left|P_\forall\left(\overline{S}_{TRA}\right)\right|}{|P_\Lambda|}, \\
R_\exists\left(\overline{S}_{CAR}\right) &= \frac{\left|P_\exists\left(\overline{S}_{CAR}\right)\right|}{|P_\Lambda|}, \\
R_\forall\left(\overline{S}_{CAR}\right) &= \frac{\left|P_\forall\left(\overline{S}_{CAR}\right)\right|}{|P_\Lambda|}.
\end{aligned} \quad (15)
$$

Table 2 lists these ratios over the 12 networks.

2.4. Wilcoxon Signed-Ranks Test.

The Wilcoxon signed-ranks test is a nonparametric statistical hypothesis test used to check whether two methods perform equally well over multiple networks [38, 39]. Let d_i be the difference in performance scores of two link prediction methods on the ith network. The differences are ranked in accordance with their absolute values; in case of ties, average ranks are assigned. Let R^+ be the sum of ranks for the networks on which the second method outperformed the first, and R^- the sum of ranks for the opposite. For a larger number of networks, the statistics

$$z = \frac{T - (1/4)N(N+1)}{\sqrt{(1/24)N(N+1)(2N+1)}} \quad (16)$$

is distributed approximately normally [39]. In (16), $T = \min(R^+, R^-)$ and N is the number of networks.

With $\alpha = 0.05$, if z is small than -1.96, we reject the null-hypothesis, which states that both methods perform equally well.

TABLE 1: The basic structural features of the giant components of the 12 networks. $|V|$ and $|E|$ are the total numbers of nodes and edges, respectively. D denotes the density, which is $D = 2|E|/|V|(|V| - 1)$. $\langle k \rangle$ and $\langle d \rangle$ present the average degree and the average shortest distance, respectively. C and r indicate the clustering coefficient [26] and assortative coefficient [36], respectively. H is the degree heterogeneity [6], defined as $H = \langle k^2 \rangle / \langle k \rangle^2$, and e is the network efficiency [37].

Networks	$|V|$	$|E|$	D	$\langle k \rangle$	$\langle d \rangle$	C	r	H	e
ADV	5042	39227	3.1E-03	15.560	3.275	0.253	-0.096	5.303	0.324
CE	297	2148	4.9E-02	14.465	2.455	0.292	-0.163	1.801	0.445
Dolphin	62	159	8.4E-02	5.129	3.357	0.259	-0.044	1.327	0.379
Email	1133	5451	8.5E-03	9.622	3.606	0.220	0.078	1.942	0.300
FW	128	2075	2.6E-01	32.422	1.776	0.335	-0.112	1.237	0.622
Hamster	1788	12476	7.8E-03	13.955	3.453	0.143	-0.089	3.264	0.317
HEP	5835	13815	8.1E-04	4.735	7.026	0.506	0.185	1.926	0.155
Karate	34	78	1.4E-01	4.588	2.408	0.571	-0.476	1.693	0.492
PB	1222	16714	2.2E-02	27.355	2.738	0.320	-0.221	2.971	0.398
USAir	332	2126	3.9E-02	12.807	2.738	0.625	-0.208	3.464	0.406
Word	112	425	6.8E-02	7.589	2.536	0.173	-0.129	1.815	0.442
Yeast	2224	6609	2.7E-03	5.943	4.376	0.138	-0.105	2.803	0.246

TABLE 2: The ratios of various seed pairs over the 12 networks.

Networks	$R_{\exists}(\overline{S}_{\triangle})$	$R_{\forall}(\overline{S}_{\triangle})$	$R_{\exists}(\overline{S}_{CAR})$	$R_{\forall}(\overline{S}_{CAR})$	$R_{\exists}(\overline{S}_{TRA})$	$R_{\forall}(\overline{S}_{TRA})$
ADV	0.001	0.001	0.807	0.750	0.018	0.014
CE	0.001	0.000	0.768	0.670	0.012	0.008
Dolphin	0.020	0.011	0.857	0.817	0.089	0.069
Email	0.008	0.005	0.881	0.841	0.070	0.057
FW	0.000	0.000	0.240	0.136	0.005	0.000
Hamster	0.014	0.005	0.893	0.817	0.074	0.048
HEP	0.007	0.007	0.881	0.876	0.029	0.028
Karate	0.004	0.000	0.743	0.706	0.034	0.011
PB	0.001	0.000	0.577	0.497	0.010	0.007
USAir	0.000	0.000	0.510	0.509	0.005	0.005
Word	0.067	0.028	0.799	0.735	0.108	0.062
Yeast	0.083	0.063	0.945	0.931	0.357	0.324

3. The Proposed Index

The link prediction problem has a familiar relationship with the network evolving mechanism [2, 40]. A recently proposed triangle growth mechanism demonstrates that various key features observed in most real-world networks can be generated in simulated networks [41]. Therefore, triangle structure information has an important effect in link formation.

In this work, we focus on a new triangle structure, namely *TRA-triangle*. A TRA-triangle passes through one seed node, one common neighbor, and one other node. In our opinion, the common neighbors that can form TRA-triangles are more important than others. Given two nodes u and v, we denote the number of triangles passing through them as $\triangle(u, v)$, which is

$$\triangle(u, v) = \begin{cases} CN(u, v), & \text{if } (u, v) \in E \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

For the example network in Figure 1(a), the triangles used for seed nodes a, b are shown in Figure 1(d). Clearly,

$\triangle(a, c) = 2$ and $\triangle(a, d) = 1$. Thus, node c is in more close contact with a than d. Given seed nodes x and y, z is one of their common neighbors. Function $\triangle(x, y; z)$ sums up the number of TRA-triangles formed by x, z, and y, z, which is

$$\triangle(x, y; z) = \triangle(x, z) + \triangle(y, z). \tag{18}$$

In this paper, we propose a new similarity index, by combining the aforementioned triangle structure and the idea of RA index [13]. For the convenience of statement, we name our new method *TRA* index. Its definition is

$$TRA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1 + \triangle(x, y; z)/2}{k_z}. \tag{19}$$

In (19), the numerator is $1 + \triangle(x, y; z)/2$. Therefore, the TRA index does not miss the effect of any common neighbor. If all common neighbors are zero-triangle-neighbors, TRA degenerates to RA. For the example network in Figure 1(a), $TRA(a, b) = (1+3/2)/4+(1+2/2)/3+(1+0/2)/2+(1+0/2)/4 = 49/24$.

TABLE 3: The AUC of different methods in 12 networks. The results are the average of 50 independent implementations with $|E_{ts}|/|E| = 0.1$. The best performance for each network is emphasized by boldface.

	CN	AA	RA	ADP	CAR	CAA	CRA	CCLP	TRA
ADV	0.8992	0.9026	0.9030	0.9033	0.8054	0.8051	0.8063	0.9011	**0.9043**
CE	0.8450	0.8613	0.8662	0.8654	0.7657	0.7677	0.7704	0.8625	**0.8713**
Dolphin	0.7832	0.7863	0.7854	**0.7866**	0.6475	0.6473	0.6473	0.7804	0.7850
Email	0.8471	0.8491	0.8488	0.8491	0.6994	0.6995	0.6996	0.8452	**0.8493**
FW	0.6053	0.6071	0.6114	0.6097	0.6192	0.6271	0.6321	0.6323	**0.6890**
Hamster	0.8037	0.8067	0.8074	0.8067	0.6542	0.6542	0.6543	0.8075	**0.8127**
HEP	0.8984	**0.8987**	**0.8987**	**0.8987**	0.7079	0.7079	0.7079	0.8624	0.8985
Karate	0.6985	0.7409	0.7523	0.7532	0.5848	0.5881	0.5880	0.7085	**0.7755**
PB	0.9192	0.9226	0.9239	0.9242	0.8926	0.8929	0.8946	0.9217	**0.9282**
USAir	0.9357	0.9466	0.9522	**0.9523**	0.9136	0.9158	0.9202	0.9391	0.9452
Word	0.6656	0.6649	0.6621	0.6651	0.5717	0.5713	0.5714	0.6727	**0.6809**
Yeast	0.7041	0.7047	0.7045	0.7047	0.5994	0.5994	0.5994	0.6972	**0.7054**

4. Experimental Results

Table 3 lists the predicted results of different methods in terms of AUC on the 12 networks. The results are obtained by averaging over 50 independent realizations for each network with testing set containing 10% links. The highest AUC value for each network is highlighted in boldface. Clearly, TRA index gets nine best results over the 12 networks. Meanwhile, TRA index outperforms the CAR, CAA, CRA, and CCLP indexes on all networks. We can see from Table 2 that, on most of the networks, there exist varying degrees of such seed node pairs with common neighbors that belong to $P_{\exists}(\overline{S}_{\triangle})$ and/or $P_{\forall}(\overline{S}_{\triangle})$. As stated in Introduction, CCLP index will give lower or zero similarity scores to those pairs. Furthermore, both values of $R_{\exists}(\overline{S}_{CAR})$ and $R_{\forall}(\overline{S}_{CAR})$ are very high on most of the networks. Particularly, on Dolphin, Email, Hamster, HEP, and Yeast, the corresponding values of $R_{\forall}(\overline{S}_{CAR})$ are greater than 0.8. This phenomenon indicates that only a very small fraction of seed node pairs with common neighbors on those networks can be assigned similarity scores by CAR-based indexes. Although there are some seed node pairs belonging to $P_{\exists}(\overline{S}_{TRA})$ and/or $P_{\forall}(\overline{S}_{TRA})$, TRA index still can assign reasonable similarity scores to them. Therefore, the results of TRA index in Table 3 are better than them of CAR, CAA, CRA, and CCLP indexes. For CN, AA, RA, and ADP indexes, ADP index performs the best, since it can penalize common neighbors by automatically adapting to the network. On Dolphin, HEP, and USAir, ADP index obtains the best accuracy; the performance of our index approximates to the best. In addition, TRA index achieves much better AUC scores than others on FW and Karate. This result suggests that TRA-triangles play an important role on these two networks. From Table 1, both networks are dense ones. Roughly speaking, the probability that there exist TRA-triangle-neighbors between seed nodes on dense networks is more than on sparse ones.

To check whether the proposed index is significantly different with compared methods, we applied Wilcoxon signed-ranks test [39] based on the results in Table 3. The pairwise test results are presented in Figure 2. From the statistical point

FIGURE 2: The results of Wilcoxon signed-ranks test based on Table 3. With $\alpha = 0.05$, if $z \le -1.96$, the null-hypothesis is rejected.

of view, our index is significantly better than others except ADP index, because ADP index has the capability of adapting to the structure of a network automatically. Although there is no statistical difference between our index and ADP index according to Wilcoxon signed-ranks test, our index performs better than ADP index in terms of AUC.

Figure 3 exhibits the changes of AUC on 12 networks when the proportion of E_{ts} in E increases from 10% to 20%. It is quite evident from Figure 3 that the AUC values of all indexes show downward trends when the proportion increases from 10% to 20% except on FW. The reason is that the increase of E_{ts} will decrease the size of training set E_{tr} and then will result in the number of common neighbors between seed nodes becoming small. Consequently, the difficulty of link prediction will enhance. The FW network, which possesses high average degree, small average shortest distance, and small-degree heterogeneity, is a very dense

FIGURE 3: The changes of AUC when $|E_{ts}|/|E|$ increases from 10% to 20% on 12 networks. Each point is obtained by averaging over 50 independent realizations.

network. Therefore, the decrease of training set gives slight influence of accuracy on FW. In addition, we can observe from Figure 3 that the performance presented by all indexes on ADV, CE, Dolphin, Email, Hamster, HEP, Karate, Word, and Yeast is very similar. On these nine networks, the AUC values of CAR-based indexes are obvious lower than those of others. On the network of FW, the results of CAR-based indexes are better than those of CN, AA, RA, and ADP indexes, because FW is a very dense network in which the ratio of CAR-triangle-neighbor is very high (see Table 2). On PB and USAir, the performance of CAR-based indexes is not as bad as on other nine networks. The reason is both networks have high average degrees, small average shortest distances, and high ratio of CAR-triangle-neighbors.

Furthermore, we list the AUC values of different methods on the 12 networks when $|E_{ts}|/|E| = 0.2$ in Table 4. The results of our index outperform others on eight among the

12 networks, while CCLP index achieves the highest value on CE.

Table 5 gives the results in terms of ranking score. These results are similar to those in Table 3. The ranking score of TRA index outperforms others except on Dolphin, HEP, and USAir. The pairwise Wilcoxon signed-ranks test results are shown in Figure 4. Similar to the test in Figure 2, TRA index is significantly better than compared methods except ADP index. As depicted above, ADP has the adaptive capability and hence performs better than other compared methods.

Figure 5 describes the changes of ranking score on 12 networks when $|E_{ts}|/|E|$ increases from 10% to 20%. Clearly, all indexes yield higher ranking scores with the increase of E_{ts}. Do not forget that higher ranking score means lower accuracy. As analyzed above, FW is very dense. Thus, the changes of AUC on FW are very slight (see Figure 3). However, the changes of ranking score on FW are more
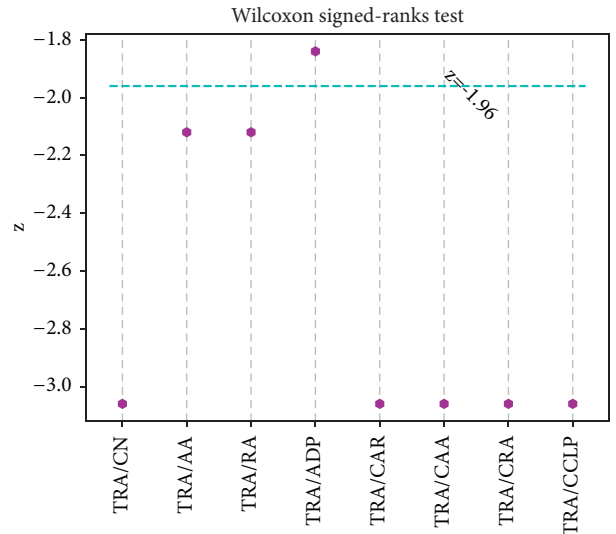
FIGURE 4: The results of Wilcoxon signed-ranks test based on Table 5. With $\alpha = 0.05$, if $z <= -1.96$, the null-hypothesis is rejected.
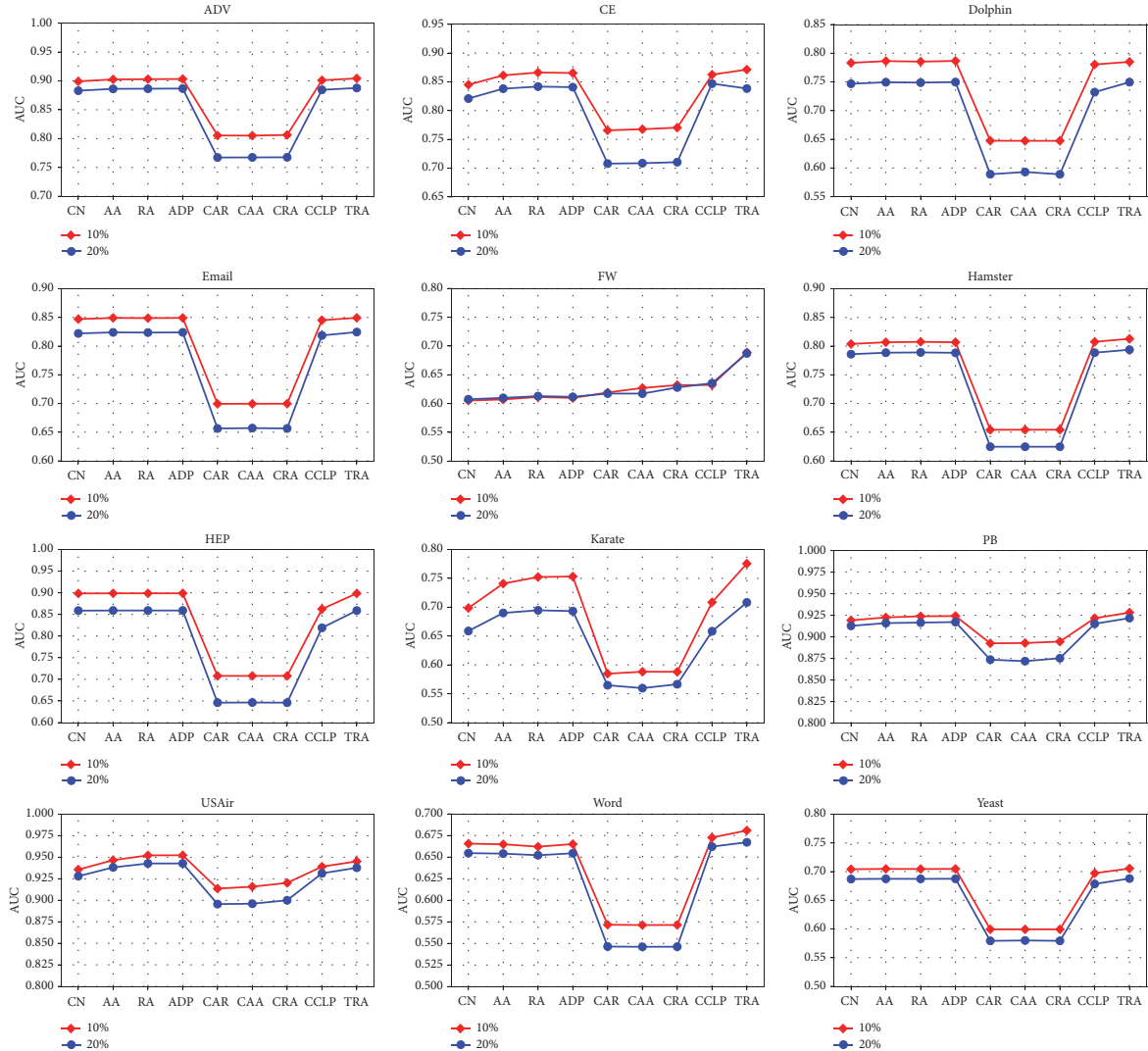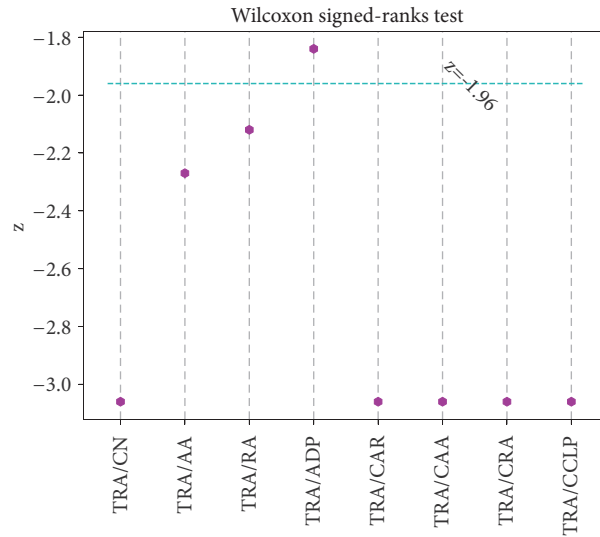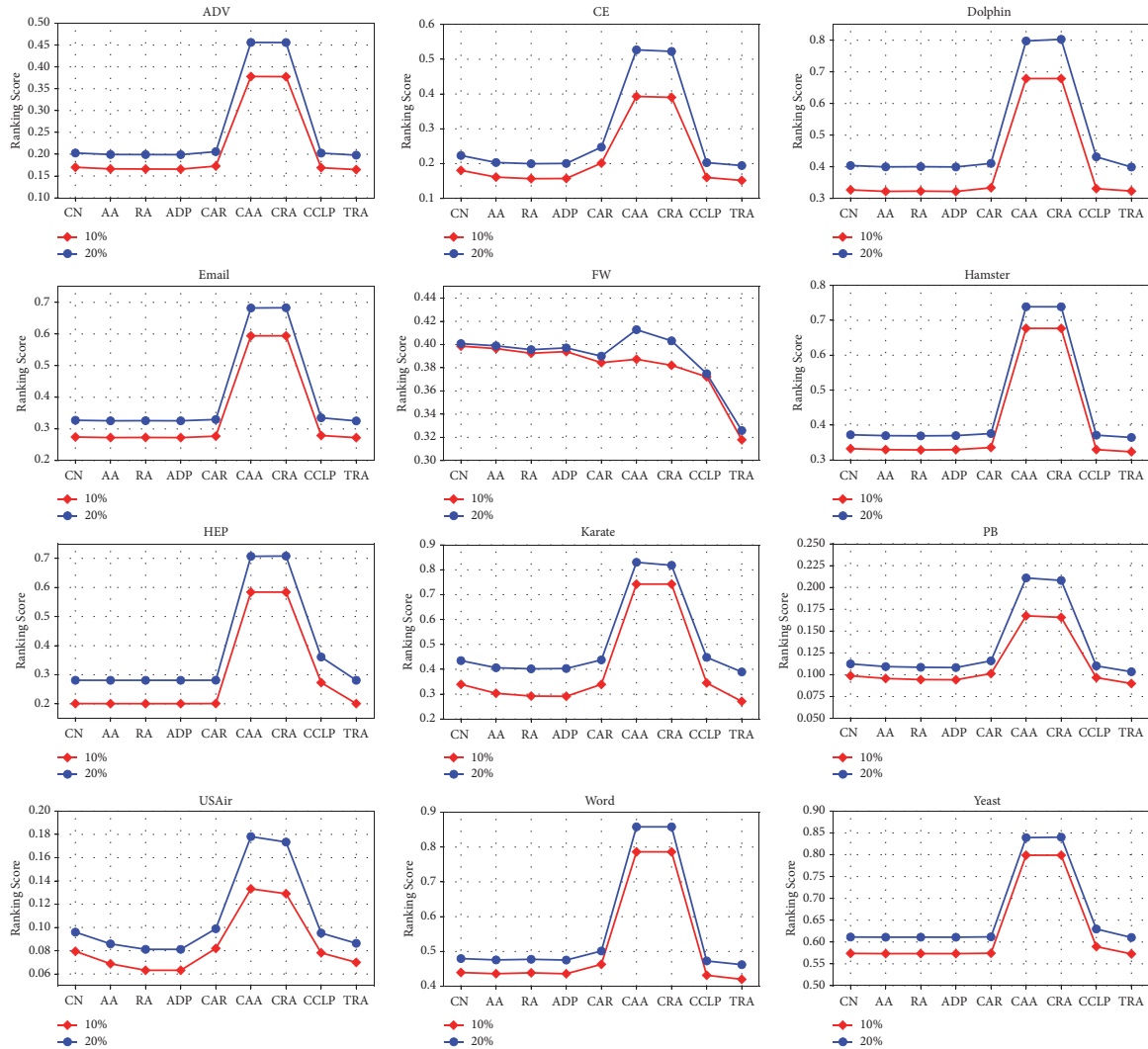

FIGURE 5: The changes of ranking score when $|E_{ts}|/|E|$ increases from 10% to 20% on 12 networks. Each point is obtained by averaging over 50 independent realizations.

TABLE 4: The AUC of different methods in 12 networks. The results are the average of 50 independent implementations with $|E_{ts}|/|E| = 0.2$. The best performance for each network is emphasized by boldface.

	CN	AA	RA	ADP	CAR	CAA	CRA	CCLP	TRA
ADV	0.8830	0.8862	0.8864	0.8867	0.7672	0.7674	0.7676	0.8845	**0.8877**
CE	0.8210	0.8381	0.8418	0.8407	0.7079	0.7086	0.7104	**0.8469**	0.8384
Dolphin	0.7468	0.7494	0.7488	**0.7497**	0.5891	0.5930	0.5890	0.7322	0.7496
Email	0.8221	0.8239	0.8236	0.8239	0.6564	0.6571	0.6565	0.8186	**0.8244**
FW	0.6075	0.6099	0.6129	0.6117	0.6174	0.6173	0.6280	0.6354	**0.6872**
Hamster	0.7859	0.7885	0.7890	0.7883	0.6246	0.6246	0.6246	0.7885	**0.7937**
HEP	0.8587	**0.8590**	**0.8590**	**0.8590**	0.6460	0.6464	0.6460	0.8190	0.8589
Karate	0.6587	0.6900	0.6946	0.6932	0.5647	0.5597	0.5664	0.6582	**0.7082**
PB	0.9128	0.9160	0.9166	0.9172	0.8736	0.8719	0.8753	0.9152	**0.9218**
USAir	0.9280	0.9382	**0.9428**	**0.9428**	0.8956	0.8960	0.9000	0.9313	0.9378
Word	0.6546	0.6541	0.6522	0.6545	0.5464	0.5461	0.5462	0.6622	**0.6672**
Yeast	0.6870	0.6874	0.6873	0.6875	0.5793	0.5800	0.5793	0.6785	**0.6879**

TABLE 5: The ranking score of different methods in 12 networks. The results are the average of 50 independent implementations with $|E_{ts}|/|E| = 0.1$. The best performance for each network is emphasized by boldface.

	CN	AA	RA	ADP	CAR	CAA	CRA	CCLP	TRA
ADV	0.1700	0.1663	0.1660	0.1657	0.1727	0.3780	0.3776	0.1690	**0.1647**
CE	0.1807	0.1613	0.1568	0.1574	0.2015	0.3930	0.3904	0.1603	**0.1518**
Dolphin	0.3271	0.3223	0.3232	**0.3220**	0.3338	0.6788	0.6788	0.3311	0.3234
Email	0.2745	0.2727	0.2731	0.2726	0.2771	0.5941	0.5942	0.2793	**0.2724**
FW	0.3986	0.3965	0.3925	0.3939	0.3844	0.3873	0.3821	0.3722	**0.3179**
Hamster	0.3323	0.3295	0.3287	0.3295	0.3357	0.6769	0.6768	0.3297	**0.3234**
HEP	0.2008	**0.2005**	**0.2005**	**0.2005**	0.2009	0.5839	0.5839	0.2729	0.2007
Karate	0.3393	0.3034	0.2922	0.2913	0.3391	0.7424	0.7424	0.3450	**0.2708**
PB	0.0988	0.0957	0.0944	0.0942	0.1013	0.1675	0.1657	0.0967	**0.0899**
USAir	0.0795	0.0688	**0.0632**	**0.0632**	0.0820	0.1332	0.1290	0.0782	0.0700
Word	0.4396	0.4362	0.4387	0.4360	0.4631	0.7862	0.7860	0.4317	**0.4201**
Yeast	0.5739	0.5734	0.5735	0.5734	0.5743	0.7989	0.7989	0.5895	**0.5727**

evident, especially for CAA and CRA indexes. The reason is that the calculation of ranking score considers all missing links. In addition, as seen in Figure 5, CAA and CRA indexes perform worse than CAR index according to ranking score. From the definitions of these three indexes, we find that both CAA and CRA indexes can get more negative impact than CAR index from zero-triangle-neighbors.

Finally, the ranking scores of all methods on the 12 networks with $|E_{ts}|/|E| = 0.2$ are listed in Table 6. Our index outperforms all other indexes except on HEP and USAir in terms of ranking score. These results are consistent with them of AUC. In contrast with that on FW, the influence of TRA-triangles on HEP and USAir is small.

From the above results, we can conclude that TRA index is superior to CAR-based indexes and CCLP index and performs better than common-neighbor-based methods on most of networks.

5. Conclusion and Discussion

Link prediction is an important research topic of complex network analysis and has a wide range of applications in various fields. Inspired by the triangle growth mechanism in network evolving [41], this paper proposed the TRA index for link prediction. When computing the similarity between two seed nodes, the proposed index not only counts the contributions of all common neighbors but also emphasizes the importance of the neighbors that can form TRA-triangles. To some extent, TRA-triangles reflect the close relationships between neighbors and seed nodes. In addition, the proposed index also adopts the theory of resource allocation [13] due to its effectiveness.

The accuracy of the TRA index is experimentally evaluated over 12 real-world networks from various fields in terms of AUC and ranking score. The experimental results show that the proposed index performs far better than CAR-based indexes. Meanwhile, our index outperforms the CCLP index because of the superior strategy in our index. For common-neighbor-based methods, the proposed index yields some improvements of accuracy on most of networks. These results indicate that combining the information of TRA-triangles and the theory of resource allocation in similarity index is a helpful idea for link prediction.

Table 6: The ranking score of different methods in 12 networks. The results are the average of 50 independent implementations with $|E_{ts}|/|E|$ = 0.2. The best performance for each network is emphasized by boldface.

	CN	AA	RA	ADP	CAR	CAA	CRA	CCLP	TRA
ADV	0.2027	0.1993	0.1991	0.1989	0.2058	0.4561	0.4558	0.2024	**0.1978**
CE	0.2234	0.2033	0.1998	0.2006	0.2473	0.5268	0.5224	0.2029	**0.1945**
Dolphin	0.4040	0.3998	0.4004	0.3995	0.4108	0.7976	0.8029	0.4315	**0.3993**
Email	0.3274	0.3257	0.3261	0.3257	0.3298	0.6821	0.6829	0.3352	**0.3253**
FW	0.4008	0.3989	0.3956	0.3971	0.3900	0.4128	0.4032	0.3749	**0.3259**
Hamster	0.3721	0.3696	0.3690	0.3697	0.3756	0.7388	0.7387	0.3710	**0.3644**
HEP	0.2810	**0.2808**	**0.2808**	**0.2808**	0.2811	0.7070	0.7079	0.3609	0.2809
Karate	0.4347	0.4061	0.4017	0.4030	0.4376	0.8301	0.8181	0.4478	**0.3893**
PB	0.1124	0.1093	0.1085	0.1082	0.1159	0.2109	0.2080	0.1101	**0.1034**
USAir	0.0960	0.0859	**0.0812**	**0.0812**	0.0989	0.1781	0.1734	0.0952	0.0864
Word	0.4796	0.4757	0.4775	0.4753	0.5013	0.8580	0.8580	0.4728	**0.4621**
Yeast	0.6114	0.6110	0.6111	0.6110	0.6118	0.8392	0.8402	0.6298	**0.6104**

There are some improved studies for our index in future. One of them is to analyze the degree of influence of TRA-triangles on different networks and further to be adaptive to set the weight of TRA-triangles on different networks. The second is to study the application of TRA index on other topics, such as community detection and anomaly detection. In addition, for learning-based link prediction approaches, TRA index can be used as a feature for a node pair.

Data Availability

The networks used in this study are available from http://deim.urv.cat/~alexandre.arenas/data/welcome.htm，http://www-personal.umich.edu/~mejn/netdata/，http://vlado.fmf.uni-lj.si/pub/networks/data/，http://noesis.ikor.org/datasets/link-prediction, and http://konect.uni-koblenz.de/networks/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Q.-M. Zhang, L. Lü, W.-Q. Wang, Y.-X. Zhu, and T. Zhou, "Potential theory for directed networks," *PLoS ONE*, vol. 8, no. 2, Article ID e55437, 2013.

[2] Q. Zhang, X. Xu, Y. Zhu, and T. Zhou, "Measuring multiple evolution mechanisms of complex networks," *Scientific Reports*, vol. 5, no. 1, 2015.

[3] L. Lü, M. Medo, C. H. Yeung, Y. Zhang, Z. Zhang, and T. Zhou, "Recommender systems," *Physics Reports*, vol. 519, no. 1, pp. 1–49, 2012.

[4] R. Guimerà and M. Sales-Pardo, "Missing and spurious interactions and the reconstruction of complex networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 106, no. 52, pp. 22073–22078, 2009.

[5] S. S. Bhowmick and B. S. Seah, "Clustering and Summarizing Protein-Protein Interaction Networks: A Survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 638–658, 2016.

[6] L. Lü and T. Zhou, "Link prediction in complex networks: a survey," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 6, pp. 1150–1170, 2011.

[7] L. Li, L. Qian, X. Wang, S. Luo, and X. Chen, "Accurate similarity index based on activity and connectivity of node for link prediction," *International Journal of Modern Physics B*, vol. 29, no. 17, 1550108, 15 pages, 2015.

[8] P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: the state-of-the-art," *Science China Information Sciences*, vol. 58, no. 1, pp. 1–38, 2014.

[9] V. Martínez, F. Berzal, and J.-C. Cubero, "A survey of link prediction in complex networks," *ACM Computing Surveys*, vol. 49, no. 4, pp. 69:1–69:33, 2016.

[10] C. Ahmed, A. ElKorany, and R. Bahgat, "A supervised learning approach to link prediction in Twitter," *Social Network Analysis and Mining*, vol. 6, no. 1, 2016.

[11] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the Association for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[12] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Social Networks*, vol. 25, no. 3, pp. 211–230, 2003.

[13] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," *The European Physical Journal B*, vol. 71, no. 4, pp. 623–630, 2009.

[14] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[15] G. Jeh and J. Widom, "SimRank," in *Proceedings of the the eighth ACM SIGKDD international conference*, p. 538, Edmonton, Alberta, Canada, July 2002.

[16] H. Tong, C. Faloutsos, and J. Pan, "Fast random walk with restart and its applications," in *Proceedings of the 6th International Conference on Data Mining (ICDM '06)*, pp. 613–622, December 2006.

[17] L. Lü, C.-H. Jin, and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 80, no. 4, Article ID 046122, 2009.

[18] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "Fast and accurate link prediction in social networking systems," *The Journal of Systems and Software*, vol. 85, no. 9, pp. 2119–2132, 2012.

[19] W. Liu and L. Lu, "Link prediction based on local random walk," *EPL (Europhysics Letters)*, vol. 89, no. 5, Article ID 58007, 2010.

[20] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi, "From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks," *Scientific Reports*, vol. 3, article 1613, no. 4, 2013.

[21] B. Chen and L. Chen, "A link prediction algorithm based on ant colony optimization," *Applied Intelligence*, vol. 41, no. 3, pp. 694–708, 2014.

[22] D. Caiyan, L. Chen, and B. Li, "Link prediction in complex network based on modularity," *Soft Computing*, vol. 21, no. 15, pp. 4197–4214, 2017.

[23] V. Martnez, F. Berzal, and J.-C. Cubero, "Adaptive degree penalization for link prediction," *Journal of Computational Science*, vol. 13, pp. 1–9, 2016.

[24] Z. Wu, Y. Lin, J. Wang, and S. Gregory, "Link prediction with node clustering coefficient," *Physica A: Statistical Mechanics and its Applications*, vol. 452, pp. 1–8, 2016.

[25] P. Massa, M. Salvetti, and D. Tomasoni, "Bowling alone and trust decline in social network sites," in *Proceedings of the 8th IEEE International Symposium on Dependable, Autonomic and Secure Computing, DASC 2009*, pp. 658–663, China, December 2009.

[26] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[27] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, "The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: can geographic isolation explain this unique trait?" *Behavioral Ecology and Sociobiology*, vol. 54, no. 4, pp. 396–405, 2003.

[28] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 6, Article ID 065103, 2003.

[29] R. E. Ulanowicz and D. L. DeAngelis, "Network analysis of trophic dynamics in south florida ecosystems," in *US Geological Survey Program on the South Florida Ecosystem*, vol. 114, 45 edition, 2005.

[30] J. Kunegis, "KONECT—the koblenz network collection," in *Proceedings of the 22nd International Conference on World Wide Web (WWW '13)*, pp. 1343–1350, May 2013.

[31] M. E. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 98, no. 2, pp. 404–409, 2001.

[32] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[33] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 U.S. Election: Divided they blog," in *Proceedings of the 3rd International Workshop on Link Discovery (LinkKDD '05)*, pp. 36–43, ACM, 2005.

[34] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 3, Article ID 036104, 19 pages, 2006.

[35] D. Bu, Y. Zhao, L. Cai et al., "Topological structure analysis of the protein-protein interaction network in budding yeast," *Nucleic Acids Research*, vol. 31, no. 9, pp. 2443–2450, 2003.

[36] M. E. Newman, "Mixing patterns in networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 67, no. 2, 2003.

[37] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, Article ID 198701, 2001.

[38] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[39] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[40] W.-Q. Wang, Q.-M. Zhang, and T. Zhou, "Evaluating network models: a likelihood analysis," *EPL (Europhysics Letters)*, vol. 98, no. 2, Article ID 28004, 2012.

[41] Z. Wu, G. Menichetti, C. Rahmede, and G. Bianconi, "Emergent complex network geometry," *Scientific Reports*, vol. 5, 2015.

Research Article

Establishment and Analysis of the Supernetwork Model for Nanjing Metro Transportation System

Yu Wei (ID) **and Sun Ning**

College of Automobile and Transport Engineering, Nanjing Forestry University, Nanjing 210037, China

Correspondence should be addressed to Yu Wei; yuweicar@163.com

In recent years, many researchers have applied complex network theory to urban public transport network to construct complex network and analyze its network performance. The original analysis method generally uses the Space L and Space R model to establish a simple link between public sites but ignores the organic link between the overall network system and the line subsystem. As an important part of urban public transport system, subway plays an important role in alleviating traffic pressure. In this paper, a supernetwork model of Nanjing metro network is established by using the supernetwork method. Three parameters, node-hyperedge degree, hyperedge-node degree, and hyperedge degree, are proposed to describe the model. The model is compared with the traditional Space L and Space P models. The study on the supernetwork model of Nanjing metro complex network shows that the network density, network centrality, and network clustering coefficient are large, and the average network distance is small, which meets the requirements of traffic planning and design. In this study, the subway line is considered as a subsystem and further simplified as a node, so that the complex network analysis method can be applied to the new supernetwork model, expanding the thinking of complex network research.

1. Introduction

With the rapid development of urban construction, more and more cities in China have opened the subway. As an important part of urban public transport system, subway plays an important role in relieving traffic pressure. Urban public transport system can be abstracted as complex network composed of stations and routes. The study of subway network is helpful to understand the evolution mechanism of public transport system and solve the problem of urban congestion.

Complex networks are characterized by complex structure and huge number of nodes. Watts and Strogatz were first proposed to have small world characteristics for complex networks [1]. Barabasi and Albert proposed scale-free power-law distribution properties [2]. Complex networks are applied to the construction and analysis of public transport network. An X L takes the bus line as the network node and uses Space R method to establish a multiweight bus road network model. By changing the different weights, the balance of the whole public transport network system is discussed [3]. From the static point of view of network topology, Bona A A D uses complex network theory to analyze the structure of public transport system in Curitiba, and compares it with the structure of public transport system in three big cities of China, including Shanghai, Beijing, and Guangzhou [4]. Manitz J proposes two methods for the cause estimation of delay in public transport networks. The application of the two methods in simulation research and in German railway system is examined [5].

Ouyang M takes China Railway System as an example, chooses three typical models based on complex network, and analyzes railway accessibility and virtual users based on traffic [6]. Zhang L has established a complex network of Jinan public transport lines by using the Space R method. It is found that the network has small world characteristics and large average clustering coefficient [7]. Mohmand Y T studied the structural characteristics of the Pakistani railway network, whose complex network shows the properties of the small world [8]. These studies include constructing complex networks of public transport networks and analyzing their structure and performance.

Complex network theory is also used to analyze the urban subway network. Based on the complex network theory, Ding

R explores the evolution process of Kuala Lumpur public rail transit network and evaluates the network performance changes in the face of different attack strategies [9]. Based on the trip data and operation schedule of Beijing subway system, Yang Y proposed a multilayer model to analyze the traffic flow pattern of subway network [10]. Feng J establishes a multilayer model of the workday and weekend flow distribution of the subway network based on the Beijing subway trip data and operation schedule [11].

Wu X established six metro complex networks in Beijing, London, Paris, Hong Kong, Tokyo, and New York and evaluated their topological efficiency and robustness [12]. Cats O establishes an evaluation model of public transport robustness and applies the model to Amsterdam urban rail transit network and evaluates the robustness of the network [13].Zhang J analyzed the complex network characteristics of the subway network in Beijing, Guangzhou, and Shanghai and studied the vulnerability of the subway network [14]. These studies include the construction of complex urban subway network, the analysis of its structural performance and robustness, and the establishment of subway network traffic flow model.

Supernetwork theory has been applied to various industries. Wang J P presents an improved hypernetwork model of knowledge diffusion algorithm and analyzes the performance of knowledge diffusion [15]. Zhao L constructs the knowledge supernetwork model of business incubators and studies the performance differences of knowledge services of different incubators by simulation [16]. Suo Q applies hypernetwork method to analyze user ratings in social networks and puts forward suggestions for collaboration in hypernetworks [17]. Cheng Q proposes a new method to reveal the community of supernetworks, which transforms the problem of community detection into the problem of DOT partitioning [18].

Wang F, taking WeChat as a sample, proposes an attractive and node-age inhomogeneous hypernetwork model [19]. Cheng Y puts forward the concept of supply-demand matching hypernetwork for manufacturing services in SOM system and reveals the matching relationship between each service and each task [20]. Lv T proposes a three-tier petroleum emergency dispatching network based on supernetwork model, which enhances regional emergency correlation by adding transfer management process [21]. Yamada T proposes a discrete network design problem based on supernetwork optimization of freight network [22]. The application of supernetwork analysis is focused on knowledge propagation model, community network analysis, and supply chain management.

Supernetwork analysis can also be seen in the subway network. Du W J puts forward a supernetwork model of urban public transport composed of conventional public transport network and urban rail transit network. Based on the external synchronization theory of coupled complex networks, the synchronization problem of urban public transport supernetwork model is studied [23]. Suo Q takes station representation as node and line representation as superedge. This paper presents a supernetwork model to describe the evolution mechanism of high-speed railway system [24]. At present, the analysis of supernetwork in metro network is limited to abstract network and simulation research, and the important parameters of metro supernetwork are not put forward and have not been applied to actual cases.

The complex network theory is used to model and analyze the urban subway network, generally using the Space L and Spacer methods. Both methods use subway stations as nodes, the Space L method establishes the links between adjacent stations on different lines, and the Space R method establishes the links between stations on the same line. The complex network model of urban subway is composed of the links between stations, and then the performance of the complex network can be analyzed. Once these two models are established, the relationship between the station and the line is neglected, and the change of the relationship between the stations is simply analyzed.

The supernetwork model can make up for this deficiency. Urban subway supernetwork is composed of main system and subsystem. The stations between the main systems are connected by some rules, which reflects the overall structure and performance of urban subway network. At the same time, the nodes in each subsystem are connected according to some rules, reflecting the connection between the lines and stations. When analyzing the performance of the main network of the metro supernetwork, the control rules of the line subsystem must be taken into account. In some cases, the line subsystem can be further simplified as a supernode, thus reflecting the overall relationship between lines. As an upgraded version of complex network, supernetwork can more effectively reflect the real structure and performance of urban subway network.

In recent years, Nanjing's public transport system has developed rapidly, opening a number of subway lines, more and more subway lines are also under construction. As an important part of public transport network system, subway can not only save traffic resources, but also provide a strong guarantee for the convenience of passengers. In this paper, the complex network of Nanjing metro is selected as the object. On the basis of the traditional complex network model Space L and Space P, the topology model of the supernetwork is constructed. In this paper, the subway supernetwork is described with the parameters of node degree, node-hyperedge degree, hyperedge-node degree, and hyperedge degree, and the network density, center degree, average distance, and clustering coefficient in complex network are extended to the theory of supernetwork, and the comparison with the traditional subway complex network model is made.

The subway can also be called the metro. In the general description of this article, it is called the subway. And according to the official name of Nanjing, it is called the metro. The supernetwork can also be called a hypernetwork. In the general description of this article, it is called the supernetwork. In the specific model, it is called hypernetwork, and its corresponding edge is also called hyperedge.

2. Nanjing Metro Network

The data of the Nanjing metro network mainly comes from the latest Nanjing bus line map issued by the Nanjing passenger transport management office and the latest tourism traffic
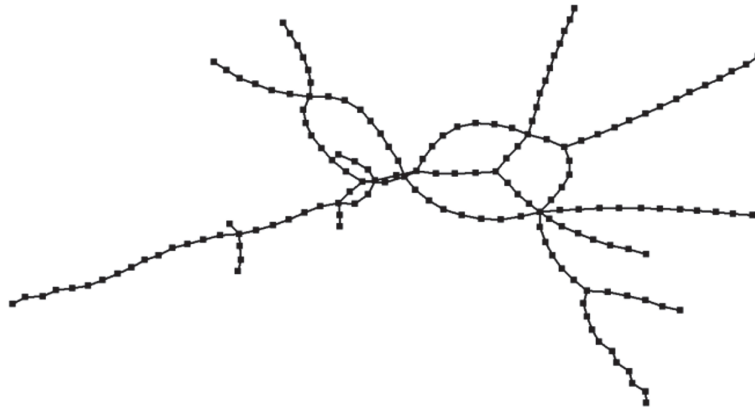
FIGURE 1: Space L spatial model of Nanjing metro complex network.
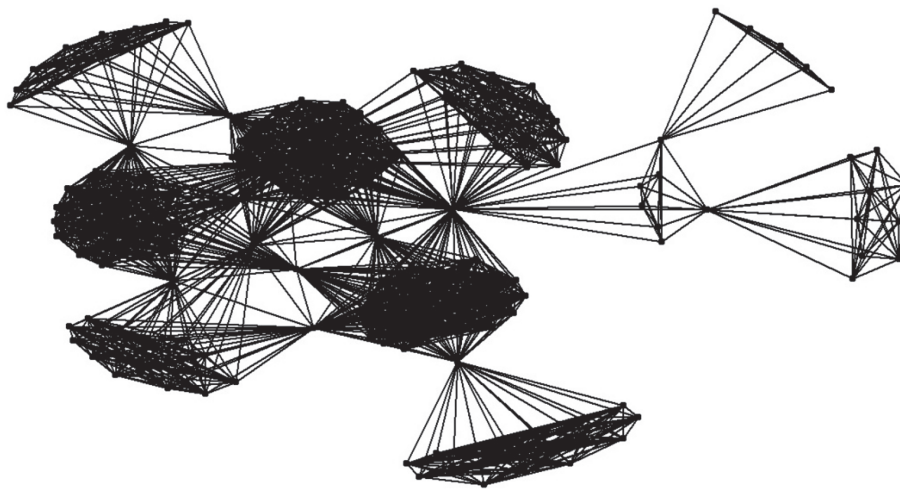
FIGURE 2: Space P spatial model of Nanjing metro complex network.

map of 2018 and the city map of Nanjing. The basic assumptions of Nanjing metro network topology are as follows.

The subway network is abstracted as an undirected network. There are differences between the upstream and downstream stations due to traffic control and other routes. Without considering the frequency of departure, the network is abstracted as a nonweighted network. The same name site is regarded as a docking site, ignoring the differences caused by the same location of individual sites but different locations. The temporary bus route diversion caused by road construction or other reasons, the cancellation or increase of subway stations, etc. shall not be considered.

There are two ways to describe the traditional traffic network topology: one is the Space L method, that is, the traffic site is regarded as a node, and if the two sites on a traffic line are adjacent, there is a link between them. Another is the Space P method, that is, the traffic network site as a node, if there is a direct traffic line between the two stations, they have a connection. From the definition, we can see that the network constructed by Space L method is the subnetwork constructed by Space P method [25].

Figure 1 shows the Space L spatial model of Nanjing metro complex network. It can be seen from the model that

the number of nodes in the network is not much, and the topology map is not complex. This is because the subway network in Nanjing is still in the process of construction, and there are still more lines to be opened in the future to meet the needs of the residents. The network presents an obvious star structure extending from the center to the periphery. At the core of the network is the core residential area of the city, surrounded by suburbs and further county towns.

Figure 2 shows the Space P spatial model of Nanjing metro complex network. It can be seen from the model that because the model represents all sites on the same line, there are ten distinct clustering subgraphs, which actually represent 10 subway lines. These lines are linked by important nodes.

3. Hypergraph and Supernetwork Model

The concept of hypergraph is proposed by BERGE in 1970. This is the first time that the theory of undirected hypergraph is established systematically, and the application of hypergraph theory in operational research is studied by using matroids. Nodes in a supernetwork represent a given set of networks, while edges and arcs represent a combined movement and a combination of preferences in a given set, and

FIGURE 3: Supernetwork topology map of Nanjing metro.

the supernetwork uniquely represents all the combination of mobile and preference dominated by the rules [26].

The definition of hypergraph is as follows: suppose V is a finite set.

If $e_i \neq H(i = 1, 2, \ldots, m)$,

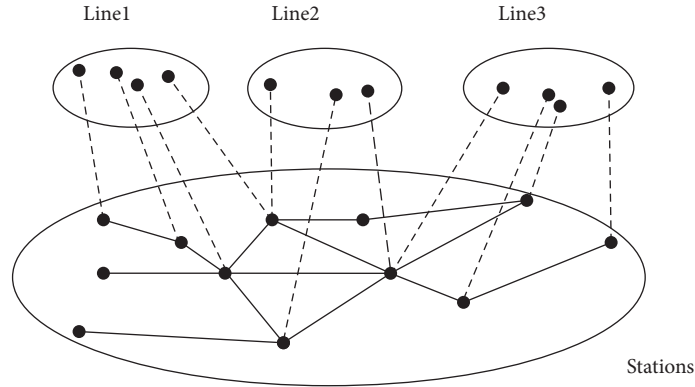$$(1) \quad \overset{m}{\underset{i=1}{Y}} e_i = V \qquad (1)$$

The two element relation $H = (V, E)$ is called a hypergraph.

The element of V, $\{v_1, v_2, \cdots, v_n\}$ is called the vertex of hypergraph, $E = \{E_1, E_2, \cdots, E_m\}$ is the edge set of hypergraph, and the set is called the edge of hypergraph.

Figure 3 shows the supernetwork topology map of Nanjing metro. The supernetwork model of the subway network consists of two parts, one is the subsystem network, which refers to the local railway lines, the other is the main system network, which refers to the overall network established between the railway stations. The main system and subsystems are independent and interrelated. Subsystem networks of the subway network include line 1, line 2, and line 3. The site on each route forms a line with certain rules. Lines, sites, and rules form the so-called hyperedge. Stations in the

main system network of a metro network are associated with certain rules, such as the Space L and Space R methods for general complex network models. However, the constraints of subsystem networks are neglected once the complex network models using these two methods are established.

The supernetwork model of metro network is different, and the organic connection between the main system network and the subsystem network is always considered. Therefore, in the analysis of the supernetwork model of subway network, the relationship between nodes-hyperedge, hyperedge-node, and hyperedge-hyperedge is included. When each subway line is simplified into a supernode, a new superedge network model is formed. Unlike Space L and Space R, each node of the hyperedge network model represents a specific subway line. A general complex network analysis method is also applicable to the superedge network model.

In the hypergraph of the supernetwork of Nanjing metro, the neighborhood matrix A reflects the relationship between the subway station and the hyperedge of the subway line. The line of the A represents the subway station, and the column of the A is the subway line. If the site belongs to a certain line, there is a relationship between the two, and the assignment is 1 or 0. A is a symmetric matrix.

$$A_{m \times n} = \begin{array}{c} \\ E_1 \\ E_2 \\ E_3 \\ E_4 \\ \vdots \\ E_m \end{array} \begin{array}{cccccc} v_1 & v_2 & v_3 & v_4 & \cdots & v_n \\ \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} \\ a_{21} & 0 & a_{23} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{32} & 0 & a_{34} & \cdots & a_{3n} \\ a_{41} & a_{42} & a_{43} & 0 & \cdots & a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & a_{m4} & \cdots & 0 \end{bmatrix} \end{array} \qquad (2)$$

where N is the number of stations on the subway network and M is the number of subway lines. $v_i(i = 1, 2, 3, 4, \cdots, n)$ stands for the subway station,. $E_j(j = 1, 2, 3, 4, \cdots, m)$

represents the subway line. $a_{ij}(i = 1, 2, 3, 4, \cdots n; j = 1, 2, 3, 4, \cdots, m)$ represents the relationship between the site and the line.
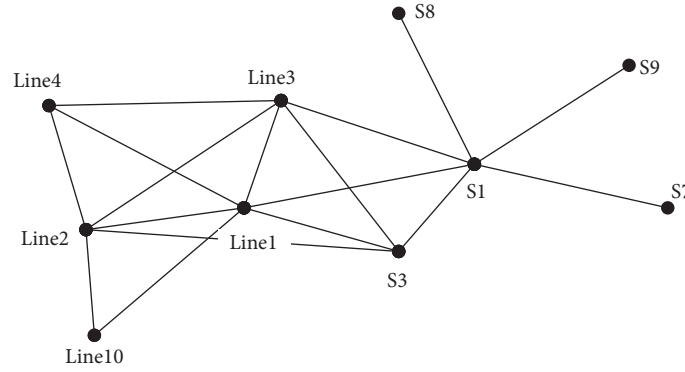
FIGURE 4: Topology of hyperedge-hyperedge relation in supernetwork of Nanjing metro.

The study in this paper further simplifies the subway supernetwork and establishes the relationship between hyperedge and hyperedge. In hypergraph $S = (E, E)$, the neighborhood matrix B reflects the relationship between the subway hyperedges. The rows and columns of B represent the metro lines. If there is the same station between the two lines, there is a relationship between them. The assignment value is 1, otherwise it is 0. B is a symmetric matrix.

$$
S_{m \times n} = \begin{array}{c} \\ E_1 \\ E_2 \\ E_3 \\ E_4 \\ \vdots \\ E_m \end{array} \begin{array}{cccccc} E_1 & E_2 & E_3 & E_4 & \cdots & E_m \\ \left[\begin{array}{cccccc} 0 & b_{12} & b_{13} & b_{14} & \cdots & b_{1m} \\ b_{21} & 0 & b_{23} & b_{24} & \cdots & b_{2m} \\ b_{31} & b_{32} & 0 & b_{34} & \cdots & b_{3m} \\ b_{41} & b_{42} & b_{43} & 0 & \cdots & b_{4m} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ b_{m1} & b_{m2} & b_{m3} & b_{m4} & \cdots & 0 \end{array}\right] \end{array} \quad (3)
$$

where M is the number of subway lines. $E_i(i = 1, 2, 3, 4, \cdots, m)$ stands for subway lines, $E_j(j = 1, 2, 3, 4, \cdots, m)$ represents subway lines, and $b_{ij}(i = 1, 2, 3, 4, \cdots, m; j = 1, 2, 3, 4, \cdots, m)$ represents the relationship between the lines.

Figure 4 shows the topology of hyperedge-hyperedge relation in supernetwork of Nanjing metro. By comparing Figure 4 with Figure 2, we can see that the model is a simplified version of the Space P spatial model. In supernetwork model, the nodes are juxtaposed. After simplification, the hyperedge space model also forms a new complex network in which the nodes represent a line, the edges of which represent a common site between the lines. The method of analyzing general complex networks is applicable to the hyperedge space model.

4. The Degree of Complex Network

4.1. Node Degree.

Those points adjacent to a point become a node's adjacent point; the number of adjacent points of a point is called the degree of the point, also known as the degree of association. The node degree is defined as the number of other nodes connected to the node. In fact, the degree of a point is also the number of lines connected to that point. If the degree of a point is 0, it is called an outlier. The node degree distribution can be described by the distribution function p (k), which indicates the probability that a randomly selected node is exactly k.

Figure 5 shows the probability distribution of node degree in Space L space of complex network in Nanjing metro. The formula is fitted to $y = 0.271x^{-1.64}$ through the data. It can be seen that the node degree distribution in the bus network Space L of Nanjing is close to the power-law distribution, which indicates that the subway network in Nanjing is a scale-free network in Space L.

According to the observation and analysis of the public transport network, the urban public transport network has the characteristics of growth and priority connectivity. Therefore, the public transport network will eventually form a scale-free network, and the distribution of the node degree in Figure 5 confirms this theory. In general, the greater the degree of a node means the more important the node is. As can be seen from Figure 5, the degree of most of the nodes is less than 6, and the node degree is basically 2. This is because the subway network structure is relatively simple and can not form a very complex network structure.

Figure 6 shows the probability distribution of node degree in Space P space of complex network in Nanjing metro. The formula is fitted to $y = 0.421x^{-0.89}$ through the data.
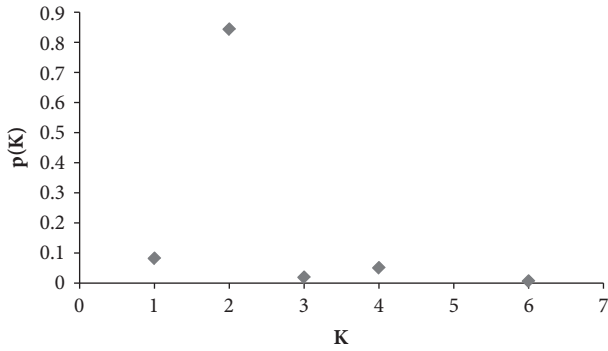
FIGURE 5: Probability distribution of node degree in Space L space of complex network in Nanjing metro.

FIGURE 6: Probability distribution of node degree in Space P space of complex network in Nanjing metro.

FIGURE 7: Probability distribution of hypernetwork node-hyperedge degree in Nanjing metro.
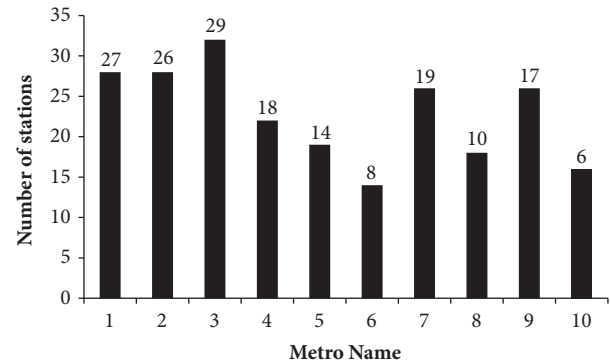
FIGURE 8: Distribution of hyperedge-node degree of hypernetwork in Nanjing metro.

It can be seen that the node degree distribution in the subway network Space P of Nanjing is close to the power-law distribution. As can be seen from Figure 6, the degree of most of the nodes is less than 60 and the degree of node concentration is between 10 and 30, which indicates the number of other sites connected by the node through the subway line.

4.2. Node-Hyperedge Degree. Node-hyperedge degree is defined as the number of hyperedges that contain the node. As shown in Figure 3, we can see that there is a node belonging to line 1 and line 2, and the node's node-hyperedge degree is 2. The node-hyperedge degree distribution can be described by the distribution function P (k), which represents the probability that the node-hyperedge degree of a randomly selected node is exactly K.

Figure 7 shows the probability distribution of hypernetwork node-hyperedge degree in Nanjing metro. The formula is fitted to $y = 0.916x^{-3.59}$ through the data. It can be seen that the probability distribution of node-hyperedge degree of Nanjing metro network is close to power-law distribution. From Figure 7, we can see that node-hyperedge degree is actually 1, 2, and 4. The number of stations containing stations is usually 1, meaning that most subway stations only have one route to go through. A few subway stations, as important transfer sites, have two routes to go through. This

is determined by the nature of the subway network, the structure presents star type radiation, and the overlapping sites are few.

4.3. Hyperedge-Node Degree. The hyperedge-node degree is defined as the number of nodes contained in a superedge. In subway hypernetwork, this parameter represents the number of subway stations contained in a line.

Figure 8 shows the distribution of hyperedge-node degree of hypernetwork in Nanjing metro. The name of the subway is 1 to 10, representing Nanjing metro 1, 2, 3, 4, 10, and suburban railway lines s1, s3, s7, s8, and s9. The degree of hyperedge-node is between 6 and 29. The subway lines in the center of the city usually have more stations, and the distance between stations is shorter, which effectively meets the needs of the residents in the central area. The suburban subway lines have fewer stations, and the distance between stations is longer, connecting the suburbs, remote counties, and airports.

4.4. Hyperedge Degree. The hyperedge degree refers to the number of other hyperedges adjacent to the hyperedge, that is, the number of other hyperedges that have common nodes with the hyperedge. In subway hyperedge network, this parameter represents the number of other subway lines connected by a subway line.

TABLE 1: Degree distribution of main sites in Nanjing metro.

Space	Space L		Space P		Node-hyperedge	
Ranking	Node	Degree	Node	Degree	Node	Degree
1	Nanjing south railway station	6	Nanjing south railway station	78	Nanjing south railway station	4
2	Yuantong	4	NanJing Railway Station	53	Yuantong	2
3	Jimingsi	4	Daxinggong	53	Youfangqiao	2
4	NanJing Railway Station	4	Xinjiekou	51	Xinjiekou	2
5	Jinma Road	4	Jimingsi	45	Xiangyu Road South	2
6	Taifeng Road	4	Taifeng Road	44	Taifeng Road	2
7	Xinjiekou	4	Youfangqiao	43	NanJing Railway Station	2
8	Gulou	4	Gulou	43	Lukou airport	2
9	Daxinggong	4	Jinma Road	42	Jinma Road	2
10	Xiangyu Road South	3	Andemen	39	Jimingsi	2
11	Andemen	3	Yuatong	38	Gulou	2
12	Youfangqiao	3	Chengxin Road	28	Daxinggong	2
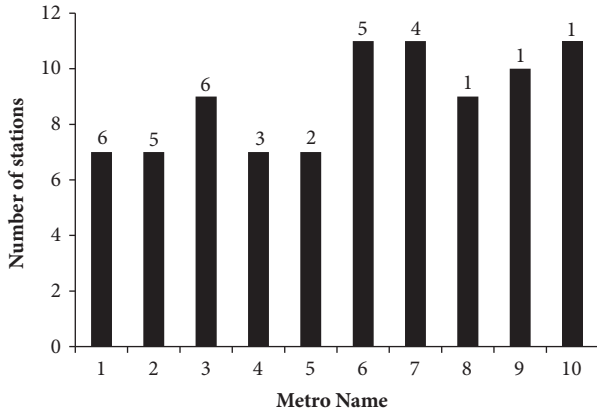
FIGURE 9: Hyperedge degree distribution of the hypernetwork in Nanjing metro.

FIGURE 10: Probability distribution of hypernetwork hyperedge degree in Nanjing metro.

Figure 9 shows the hyperedge degree distribution of the hypernetwork in Nanjing metro. The name of the subway is 1 to 10, representing Nanjing metro 1, 2, 3, 4, 10, and suburban railway lines s1, s3, s7, s8, and s9. The value of the superedge is between 1 and 6. Metro lines 1, 2, and 3 have a higher node-hyperedge degree and have a better switching function. The suburban subway s1 has a superedge of 5, because it connects the suburbs, airports, railway stations, and remote county towns. Metro line s3 has a superedge of 4, because it connects some suburban lines.

Figure 10 shows the probability distribution of hypernetwork hyperedge degree in Nanjing metro. If the hyperedge degree distribution is described by the distribution function p (k), the probability of a hyperedge of a randomly selected hyperedge is exactly k. From the probability distribution map, the hyperedge does not obey the power-law distribution, but it is similar to the two power function after fitting.

4.5. Analysis of Public Hub Sites. Table 1 shows the highest degree of 12 nodes in Space L, Space P, and node-hyperedge space. These sites, known as public pivot points, play a vital role in the urban public transport network, connected to not only a large number of subway stations, but also a number of bus stations and many of the lines through the site.

It can be seen from Table 1 that the node degree of Space P space is 6. The maximum node degree of Space P space is 78, and the range of numerical fluctuation is large. The node-hyperedge space is 2 except for one node with 4. In the three spaces, the most important is the Nanjing south railway station, which connects the suburban, railway station, and the airport's subway lines, which has played an important role in the transfer. In the three spaces, the top ranking sites are basically unchanged. These are important public hub sites.

5. Spatial Characteristics of Metro Network

In this paper, three spatial models of urban subway network are constructed by using the space L, space R and superedge

TABLE 2: Spatial characteristics of the complex network of Nanjing metro.

characteristics	Space L	Space P	Hyperedge Space
Network size	159	159	10
Network density (%)	1.31	13.69	37.78
Network centrality (%)	2.52	36.13	36.11
Network average distance	16.77	2.34	1.82
Network clustering coefficient (%)	0	95.8	67.6

space methods. The superedge network model simplifies each subway line into a supernode. If each line has the same station, the supernodes are connected. When analyzing the performance of Space L and Space R models in subway networks, the general spatial characteristic parameters include network size, network density, network center degree, network average distance, and network clustering coefficient [9, 10]. The analysis method is also applicable to the superedge network model.

Table 2 shows the spatial characteristics of Nanjing metro complex network and uses the network size, network density, network center degree, network average distance, and network clustering coefficient of five indicators. In this paper, three models of Space L, Space P, and hyperedge space are selected for comparison. The hyperedge space model is shown in Figure 4. The network size of Space L and Space P space is 159, which means that there are 159 subway stations. The size of the network in the hyperedge space is 10, which means that there are 10 subway lines.

5.1. Network Density. Network density refers to the degree of closeness between nodes in a network. Network G's network density d (G) is defined as

$$d(G) = \frac{2M}{[N(N-1)]} \tag{4}$$

where M is the number of connections actually owned in the network and N is the number of network nodes. The range of network density is [0, 1]. When the network is completely connected, the network density is 1, while the actual network density is usually much less than 1.

As can be seen from Table 2, the network density of Space L is relatively low, because the subway lines are still relatively small, and the structure presents a star-shaped loose structure. The network density of Space P is the result of the characteristics of the structure model, so the connection between the stations on each line has been established, and the density value of the network is improved. In fact, there are relatively few links between the lines. The density of network in hyperedge space is relatively high, because the model reflects the relationship between subway lines, and the distribution is more balanced in the whole region.

5.2. Network Centrality. Degree centrality is divided into node centrality and network centrality. The former refers to the degree of centrality among the nodes in which the nodes are directly connected to them, while the latter focuses on the central degree of the whole network, representing the degree

of centralization of the entire network, that is, the extent to which the entire network organizes the operation around a node or a group of nodes. The degree centrality C_D (v_i) of node v_i is defined as

$$C_D(V_i) = \frac{k_i}{N-1} \tag{5}$$

In all networks containing N nodes, assume that network $\mathbf{G}_{\text{optimal}}$ maxims the following formula:

$$H = \sum_{i=1}^{N} \left[C_D(V_{\max}) - C_D(V_i) \right] \tag{6}$$

In the formula, v_i is the node of the network $\mathbf{G}_{\text{optimal}}$, and $V_{\max}$ represents the node with the largest degree of centrality in the network $\mathbf{G}_{\text{optimal}}$.

For a network G containing N nodes, $V_{\max}$ means that it has the largest degree of centrality. Figure $\mathbf{G}_{\text{optimal}}$ for star network, the degree centrality C_D of network G is defined as

$$C_D = \frac{1}{N-2} \sum_{i=1}^{N} \left[C_D(V_{\max}) - C_D(V_i) \right] \tag{7}$$

As can be seen from Table 2, the network centrality of Space L is relatively small, which is due to loose structure; no node has a larger degree of node. The network center of Space P is relatively large, because the connection between lines is associated with all sites on different lines and thus presents better centrality. The network center of the hyperedge space is relatively large, because some important metro lines are effectively connected to other suburban metro lines, such as line 1, line 2, and line 3 of the Nanjing metro.

5.3. Network Average Distance. In mathematics, physics, and sociology, the small world network is a type of mathematical graph, in which most of the nodes are not adjacent to each other, but most of the nodes can arrive at a few steps from any other point. Small world networks are usually measured by means of two parameters: average distance and clustering coefficient. The small world standard has a small network average distance L and a high clustering coefficient C.

Distance refers to the total number of lines that a node must pass through in its path to another node, i.e., the length of the shortcut between two points. Mean distance represents the average distance between all pairs of points in a graph. The overall reachability of the network is better than the average distance, but the connectivity of the whole network cannot be truly reflected by the connected distance in the case that the

whole network is not in the connected state, but in the case of multiple subgraphs. Although many real networks have large number of nodes, the average distance is surprisingly small. This is the so-called small world effect.

For the undirected simple graph, the formula is as follows:

$$L = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} d_{ij} \qquad (8)$$

where L is the average distance of the network, N is the total number of nodes, and the distance from node i to node j.

As you can see from Table 2, the network average distance of Space L is larger because it represents the length between one site and another, and the space model's star structure determines the distance to the suburb. The average distance of Space P is 2.34, which means the average transfer is 2.34 times from one subway station to another. Considering the shortcut of the subway, the transfer efficiency is still high. The average distance in the hyperspace is 1.82, which means that it is more efficient to transfer from one subway line to another through 1.82.

5.4. Network Clustering Coefficient. According to the graph theory, the clustering coefficient is the coefficient that represents the degree of node clustering in a graph. Evidence shows that in the real network, especially in a specific network, the nodes tend to establish a set of close organizational systems because of the relative high density connection points. In real-world networks, this probability is often greater than the average probability of randomly setting up a connection between two nodes.

First of all, we look at the definition of the clustering coefficient of the nodes. If the node v_i is connected directly with the k_i node, the maximum number of possible edges between the k_i nodes for the undirected network is $k_i(k_i - 1)/2$, while the actual number of edges is M_i.

$$C = \sum_{i=1}^{n} C_i = \sum_{i=1}^{n} \frac{2M_i}{k_i(k_i - 1)} \qquad (9)$$

The network clustering coefficient C is the average clustering coefficient C_i of all nodes i. It is obvious $0 \leq C \leq 1$, where k_i represents the number of all adjacent nodes of node i and N represents the number of all nodes.

It can be seen from Table 2 that the clustering coefficient of Space L is zero because of the loose topology of space. Space P's network clustering coefficient is large because the sites on each line are set up to connect when building the model. The network clustering coefficient of the hyperedge space is relatively large, because the subway lines in some urban areas have played an important connection with the lines of the suburbs, airports, and railway stations.

6. Conclusions

When using the general complex network method to analyze the network, the nodes are often regarded as independent, ignoring the small group effect of the network. Hypergraph and hypernetwork method make up for this deficiency to a certain extent. This paper chooses Nanjing metro complex network as the research object, establishes the space L, space P, and hypernetwork model, and compares the three network structures. The hypernetwork model reflects the relationship between the subway station and the subway lines and the relationship between the subway lines. The analysis shows that the network density, network centrality, and network centrality of the metro hypernetwork in Nanjing are large, and the average distance of the network is small, which is in line with the ideal traffic planning and design.

The public transport hub sites extracted from the hypernetwork model are similar to the other two models. This paper only analyzes the complex network of Nanjing metro and can further expand to the bus system, shared bicycle system, and uses the hypernetwork model to establish a higher level, more complex system, and analyze the connection. The application of the hypernetwork model is only an undirected simple network, and the relationship established is only a subordinate relationship between the line and the site. The future hypernetwork model can be extended to a directed weighted network, to establish a more complex model to consider travel costs and travel preferences and to apply to solving other traffic problems.

Appendix

Nanjing Metro Data

The data of the Nanjing metro network mainly comes from the latest Nanjing bus line map issued by the Nanjing passenger transport management office and the latest tourism traffic map of 2018 and the city map of Nanjing. The basic assumptions of Nanjing metro network topology are as follows.

By May 2018, Nanjing metro network has opened 10 lines, 1, 2, 3, 4, and 10 and the suburban railway lines S1, S3, S7, S8, S9, which are composed of 159 subway stations. The opening order is 1, 2, 10, S1, S8, 3, 4, S3, S9, and S7.

The metro data are as follows:

Line 1: maigaoqiao, hongshandongwuyuan, nanjingzhan, xinmofanmalu, xuanwumen, gulou, zhujianglu, xinjiekou, zhangfuyuan, sanshanjie, zhonghuamen, andemen, tianlongsi, ruanjiandadao, huashenmiao, nanjingnanzhan, shuanglongdadao, hedingqiao, shengtailu, baijiahu, xiaolongwan, zhushanlu, tianyindadao, longmiandadao, nanyidajiangsujingmaoxueyuanzhan, nanjingjiaoyuan, zhongguoyaokedaxue

Line 2: youfangqiao, yurundajie, yuantong, aotidong, xinglongdajie, jiqingmendajie, yunjinlu, mochouhu, hanzhongmen, shanghailu, xinjiekou, daxinggong, xianmen, minggugong, muxuyuan, xiamafang, xiaolingwei, zhonglingjie, maqun, jinmalu, xianhemen, xuezelu, xianlinzhongxin, yangshangongyuan, nandaxianlinxiaoqu, jingtianlu

Line 3: linchang, xinghuolu, dongdachengxianxueyuan, taifenglu, tianruncheng, liuzhoudonglu, shangyuanmen,

wutangguangchang, xiaoshi, nanjingzhan, nanjinglinyedax-uexinzhuang, jimingsi, fuqiao, daxinggong, changfujie, fuzimiao, wudingmen, yuhuamen, qiazimen, daminglu, mingfaguangchang, nanjingnanzhan, hongyundadao, shengtaixilu, tianyuanxilu, jiulonghu, chengxindadao, dongdajiulong-huxiaoqu, mozhoudonglu

Line 4: longjiang, nanyiershicaochang, yunnanlu, gulou, jimingsi, jiuhuashan, gangzicun, jiangwangmiao, wangjiawan, jubaoshan, suningzongbuxuzhuang, jinmalu, huitonglu, lingshan, dongliu, mengbei, xiganghuashu, xianlinhu

Line 10: andemen, xiaoxing, zhongsheng, yuantong, aotizhongxin, mengdudajie, lüboyuan, jiangxinzhou, linjianglu, pukouwanhuicheng, nanjinggongyedaxue, longhualu, wendelu, yushanlu

S1: nanjingnanzhan, cuipingshan, fochengxilu, jiyindadao, zhengfangzhonglu, xiangyulubei, xiangyulunan, lukoujichang

S3: gaojiachong, linshan, qiaolinxincheng, shiqihe, shuanglong, lanhuatang, maluowei, liucun, tianbao, gaomiaolu, wuhoujie, pingliangdajie, yongchulu, youfangqiao, jiaxi, chunjianglu, tiexinqiao, jingmingjiayuan, nanjingnanzhan

S7: lukoujichang, jichangdong, zhetang, zhetangxinqu, jinshan, tuanshan, lishui, zhongshandonglu, jinlonglu, wuxiangshan

S8: taishanxincun, taifenglu, gaoxinkaifaqu, xinxigongchengdaxue, xiejiadian, dachang, getang, changlu, huagongyuan, liuhekaifaqu, longchi, xiongzhou, fenghuangshangongyuan, fangzhouguangchang, shenqiao, babaiqiao, jinniuhu

S9: xiangyulunan, tongshan, shiqiu, mingjue, tuanjiewei, gaochun

Data Availability

The Nanjing metro data used to support the findings of this study are included within Appendix.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[2] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[3] X. L. An, L. Zhang, Y. Z. Li et al., "Synchronization analysis of complex networks with multi-weights and its application in public traffic network," *Physica A Statistical Mechanics & Its Applications*, vol. 412, no. 10, pp. 149–156, 2014.

[4] A. A. De Bona, K. V. O. Fonseca, M. O. Rosa, R. Lüders, and M. R. B. S. Delgado, "Analysis of Public Bus Transportation of a Brazilian City Based on the Theory of Complex Networks Using

the P-Space," *Mathematical Problems in Engineering*, vol. 2016, Article ID 3898762, 12 pages, 2016.

[5] J. Manitz, J. Harbering, M. Schmidt et al., "Source estimation for propagation processes on complex networks with an application to delays in public transportation systems," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 66, no. 3, pp. 521–536, 2017.

[6] L. Zhang, J. Lu, X. Yue, J. Zhou, Y. Li, and Q. Wan, "An auxiliary optimization method for complex public transit route network based on link prediction," *Modern Physics Letters B. Condensed Matter Physics, Statistical Physics, Applied Physics*, vol. 32, no. 5, 1850066, 22 pages, 2018.

[7] Y. T. Mohmand and A. Wang, "Complex network analysis of Pakistan railways," *Discrete Dynamics in Nature and Society*, vol. 2014, Article ID 862612, 5 pages, 2014.

[8] M. Ouyang, L. Zhao, L. Hong, and Z. Pan, "Comparisons of complex network based models and real train flow model to analyze Chinese railway vulnerability," *Reliability Engineering & System Safety*, vol. 123, pp. 38–46, 2014.

[9] R. Ding, N. Ujang, H. B. Hamid, and J. Wu, "Complex network theory applied to the growth of Kuala Lumpur's public urban rail transit network," *PLoS ONE*, vol. 10, no. 10, Article ID e0139961, 2015.

[10] Y. Yang, Y. Liu, M. Zhou, F. Li, and C. Sun, "Robustness assessment of urban rail transit based on complex network theory: A case study of the Beijing Subway," *Safety Science*, vol. 79, pp. 149–162, 2015.

[11] J. Feng, X. Li, B. Mao, Q. Xu, and Y. Bai, "Weighted Complex Network Analysis of the Different Patterns of Metro Traffic Flows on Weekday and Weekend," *Discrete Dynamics in Nature and Society*, vol. 2016, no. 1, pp. 1–10, 2016.

[12] X. Wu, H. Dong, K. T. Chi et al., "Analysis of metro network performance from a complex network perspective," *Physica A Statistical Mechanics & Its Applications*, p. 492, 2017.

[13] O. Cats, G.-J. Koppenol, and M. Warnier, "Robustness assessment of link capacity reduction for complex networks: Application for public transport systems," *Reliability Engineering & System Safety*, vol. 167, pp. 544–553, 2017.

[14] J. Zhang, S. Wang, and X. Wang, "Comparison analysis on vulnerability of metro networks based on complex network," *Physica A: Statistical Mechanics and its Applications*, vol. 496, pp. 72–78, 2018.

[15] J.-P. Wang, Q. Guo, G.-Y. Yang, and J.-G. Liu, "Improved knowledge diffusion model based on the collaboration hypernetwork," *Physica A: Statistical Mechanics and its Applications*, vol. 428, pp. 250–256, 2015.

[16] L. Zhao, H. Zhang, and W. Wu, "Knowledge service decision making in business incubators based on the supernetwork model," *Physica A: Statistical Mechanics and its Applications*, vol. 479, pp. 249–264, 2017.

[17] Q. Suo, S. Sun, N. Hajli, and P. E. D. Love, "User ratings analysis in social networks through a hypernetwork method," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7317–7325, 2015.

[18] Q. Cheng, Z. Liu, J. Huang, and G. Cheng, "Community detection in hypernetwork via Density-Ordered Tree partition," *Applied Mathematics and Computation*, vol. 276, pp. 384–393, 2016.

[19] F. Wang and J. Guo, "Research on Mobile Social Non-uniform Hypernetwork Evolving model," *International Journal of Mobile Communications*, vol. 16, no. 1, p. 1, 2018.

[20] Y. Cheng, F. Tao, D. Zhao, and L. Zhang, "Modeling of manufacturing service supply–demand matching hypernetwork in service-oriented manufacturing systems," *Robotics and Computer-Integrated Manufacturing*, vol. 45, pp. 59–72, 2017.

[21] T. Lv, Y. Nie, C. Wang, and J. Gao, "Cross-regional emergency scheduling planning for petroleum based on the supernetwork model," *Petroleum Science*, vol. 15, no. 3, pp. 666–679, 2018.

[22] T. Yamada and Z. Febri, "Freight transport network design using particle swarm optimisation in supply chain-transport supernetwork equilibrium," *Transportation Research Part E: Logistics and Transportation Review*, vol. 75, pp. 164–187, 2015.

[23] W.-j. Du, J.-g. Zhang, X.-l. An, S. Qin, and J.-n. Yu, "Outer synchronization between two coupled complex networks and its application in public traffic supernetwork," *Discrete Dynamics in Nature and Society*, vol. 2016, Article ID 8920764, 8 pages, 2016.

[24] Q. Suo and J. L. Guo, "The evolutionary mechanism of high-speed railway system based on hypernetwork theory," *International Journal of Modern Physics B*, vol. 32, no. 15, Article ID 1850182, 2018.

[25] V. Latora and M. Marchiori, "Is the Boston subway a small-world network?" *Physica A: Statistical Mechanics and its Applications*, vol. 314, no. 1, pp. 109–113, 2002.

[26] E. Estrada and J. A. Rodríguez-Velázquez, "Subgraph centrality in complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, no. 5, 2005.

Research Article

Scare Behavior Diffusion Model of Health Food Safety Based on Complex Network

Jun Luo ⓘ,[1,2] **Jiepeng Wang,**[3] **Yongle Zhao,**[1] **and Tingqiang Chen** ⓘ[3]

[1]*Business School, Hohai University, Nanjing, China*
[2]*School of Health Economics and Management, Nanjing University of Chinese Medicine, Nanjing, China*
[3]*School of Economics and Management, Nanjing Tech University, Nanjing, China*

Correspondence should be addressed to Tingqiang Chen; tingqiang88888888@163.com

This study constructs a heterogeneous model of health food safety scare behavior diffusion through a complex network model by considering health food safety information transparency and health food consumers' ability to process information. This study first analyzes the effects of network structure and heterogeneity of health food consumers on the health food safety scare behavior diffusion using network stochastic dominance theory. Subsequently, a computer mathematical simulation is performed to explore the characteristics and laws of the evolution of health food safety scare behavior diffusion. The following three major conclusions can be drawn from the results. First, increases in the health food safety information transparency, the health food consumers' ability to process information, and the recovery rate of health food consumers can increase the threshold of the rate of health food safety scare behavior diffusion. The health food safety information transparency and the recovery rate of health food consumers show marginal incremental rising characteristics in relation to the rate of health food safety scare behavior diffusion, whereas the health food consumers' ability to process information shows a marginal diminishing rising characteristic in relation to the rate of health food safety scare behavior diffusion. Second, increases in the health food safety information transparency, the health food consumers' ability to process information, and the recovery rate of health food consumers can decrease the scale of the health food safety scare behavior diffusion. The health food safety information transparency shows a marginal diminishing decreasing characteristic in relation to the scale of the health food safety scare behavior diffusion, whereas the health food consumers' ability to process information and the recovery rate of the health food consumers show marginal incremental decreasing characteristics in relation to the scale of the health food safety scare behavior diffusion. Finally, the network structure of health food consumers significantly affects the health food safety scare behavior diffusion. A high heterogeneity of the health food consumer network indicates a high threshold of the rate of health food safety scare behavior diffusion and low diffusion scale.

1. Introduction

The concept of food safety is continuously developing. Understanding food safety is a dynamic development process [1–5]. Food safety includes not only food security but also food quality and health safety [3–5]. In recent years, numerous media reports on food industry emergencies have been persistent, thereby transforming a sudden food safety problem into a serious public policy and social problem. For example, incidents related to salt supply safety broke out from March 16, 2011, to March 18, 2011, under the influence of scare behavior diffusion in China after the 2011 earthquake in Japan. The food safety scare behavior caused by food safety accidents can spread through certain media to consumers in their healthy state as a product of the development and evolution of emergencies [3–6], thereby provoking panic behavior with a significant herd effect. Many food safety accidents are safe in themselves, but the loss induced by these accidents is much greater than the direct loss from the accidents themselves [7, 8]. Therefore, food safety accidents are detrimental to the healthy development of social stability and food industry [9, 10].

Currently, the party and the government in China introduced the plan "Healthy China 2030." The development of

new business forms of the health industry, as an important part of "Healthy China 2030," must be considered. The health food industry, as an emerging health industry, has numerous apparent shortcomings, including the imperfection of the industrial system, low level of the industrial system, low coordination effect, and lack of industry standards and norms [11]; consequently, the health food industry has received extensive attention from the society. Specifically, chaos may result from health food safety accidents, such as those in which health food passes off as medicine, health food contains illegal drug components, and functional efficacy of health food is inconsistent with reality. The health food safety scare behavior caused by such accidents has affected the implementation of China's strategy of building a healthy China. Furthermore, existing empirical studies show that improving the transparency of food safety supervision information can reduce the adverse effect of food safety accidents [3–5, 12–17]. Mol [16] investigated to what extent and how China's transparency institutions and practices regarding food production and products play a role in governing food quality and safety. Chen et al. [3–5] found that the transparency of food safety supervision in China is basically qualified but remains at a poor level; this transparency may be improved by establishing an index system for food safety supervision information transparency and evaluating different regulatory bodies in China. Easing the rumor-driven "herd behavior" is not conducive to promoting the healthy development of food industry. The role of food information transparency in alleviating the food safety problem has gradually attracted the attention of numerous scholars given the reinforcement of food safety management [3–5, 16, 18]. However, health food is in the primary stage of development in China. Moreover, research on information transparency in the health food industry is limited; let alone research on health food safety scare behavior with consideration of health food information transparency. Therefore, in accordance with the suggestion of Chen et al. [3–5], two key factors, namely, health food information transparency and health food consumers' ability to process information, are considered in the current study by exploring the internal mechanism and evolution law of the health food safety scare behavior diffusion among health food consumers. The results of the present study can change the current situation of health food safety in China to a certain extent and promote the formation of social cogovernance of health food safety.

Complex network theory and methods have been developed in several studies [19–24] and have been applied to various fields. A network is a collection of nodes and edges. A complex network is composed of many nodes and edges of connecting nodes. Its complexity is mainly manifested in the number and properties of the nodes and edges. It is usually a high generalization or abstraction of complex systems and phenomena. Compared with a general network, a complex network has six unique characteristics that reflect its complexity. First, a complex network has a complicated structure. The number of its nodes is large, and the network structure presents many different characteristics. Second, a complex network presents network evolution, as reflected in the emergence and disappearance of nodes or connections.

Third, a complex network has a connection diversity. The connection weights between nodes vary, and the connections may be directional. Fourth, a complex network has a dynamic complexity. A node set may belong to a nonlinear dynamic system. Fifth, a complex network features a node diversity. Nodes in complex networks can represent anything. Finally, a complex network presents multiple-complexity integration. The complexities that result from the five aforementioned characteristics influence one another, and their interplay leads to unpredictable results. In addition to these six complex characteristics, a complex network also has a dynamic complexity in time and space, and its network behavior is also complex. Many real-world complex systems, such as transportation network, the Internet, investment network, and disease contagion network, can be in the form of a complex network by abstracting, description, characterization, and analysis. The capability of complex networks to represent real-world phenomena and their dynamic evolution behaviors has prompted science researchers in various fields to describe, analyze, and model complex networks and establish algorithms thereof in theoretical and empirical studies. By the end of the 20th century, the WS network proposed by Watts and Strogatz [19] and the BA network proposed by Barabási and Albert [20] epitomize the new era of complex network research. At present, complex network theory, which is the general methodology for analyzing all types of complex systems and their complex phenomena in the real world, has infiltrated studies in many fields, including statistical physics, biological sciences, and humanistic social sciences [25–27]. Insights into and methods based on complex networks have become a research interest in the scientific study community.

The complex network constructed in this study is the health food consumer network. Currently, health food consumers in China are mainly the elderly, and the health food consumer network becomes increasingly complex while the aging population in China expands. On the one hand, health food consumers exhibit different cognitive performances in terms of health food [3–5, 28, 29]. Several health food consumers are protected from health food safety scare behavior considering their abundant knowledge of health food; these consumers can inhibit the spread of this behavior. By contrast, certain health food consumers are vulnerable to the influence of connected health food consumers, thereby showing remarkable herd behavior [30]. On the other hand, various health food consumers possess different psychological qualities or psychological cognitions. In the face of health food safety scare behavior, consumers with high psychological quality can be minimally affected or can even avoid being affected by this scare behavior. In consideration of this scenario, this study investigates health food consumers' ability to process information. Heterogeneity in such a complex network cannot be ignored. Therefore, this study uses stochastic dominant theory to analyze the health food safety scare behavior diffusion under different network structures.

The concept of stochastic dominance was first explicitly proposed by Quirk and Saposnik in 1962 [31], and they associated it with the traditional expected utility principle. Since then, stochastic dominance theory has been the basis of risk decision methods. This theory has been a widely
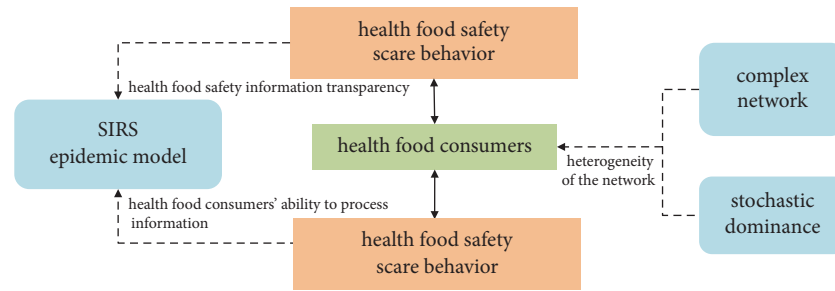
FIGURE 1: Technology roadmap of the article.

recognized decision analysis tool used by economic agents to make a decision of behavior in uncertainty cases. Stochastic dominance is a nonparametric decision analysis method in which various optional results and their corresponding objective probabilities are analyzed to filter the nondominant schemes and make risk decisions. Studies on stochastic dominance date back to Karamata [32]. However, the theory attracted the attention of the academic community only in the late 1960s and early 1970s when the studies of Hadar and Russel [33], Hanoch and Levy [34], and Rothchild and Stiglitz [35] were reported. Particularly in the 1970s, numerous studies on stochastic dominance theory and its applications emerged in the research literature. Several of these studies proposed decision-making methods, such as the third-order and expectation stochastic dominance. The most important representatives of these studies is that of Fishburn [36], who extended the concept of stochastic dominance to random variables. Jackson [37] adopted the stochastic dominance method to analyze a social network structure and investigated the influence of structural characteristics and network heterogeneity on social behavior. Bian et al. [38] used the stochastic dominance to explore the evolvement of investors' behavior in stock market. In the present study, we use the same method used by Jackson [37] to analyze the influence of structure and heterogeneity of the health food consumer network on the health food safety scare behavior diffusion.

Health food safety scare behavior diffusion is a typical proliferation problem, and its mechanism is similar to those of the spread of infectious diseases. The technology roadmap of this study is illustrated in Figure 1. We establish a network diffusion model on the basis of the SIRS model of health food safety scare behavior and analyze the influences of health food information transparency, health food consumers' ability to process information, and network structure on the mechanism of health food safety scare behavior diffusion. This study can provide a reference for controlling health food safety scare behavior diffusion and reducing its effect on the society.

The structure of this study is arranged as follows. Section 2 discusses the infectious disease principles and characteristics of health food safety scare behavior diffusion in two aspects: (1) adaptability of the epidemic model and (2) health food safety scare behavior diffusion mechanism. Section 3 presents the model constructed for health food safety scare behavior diffusion and the corresponding theoretical analysis. Section 4 presents a computer simulation

analysis of the health food safety scare behavior diffusion. Section 5 summarizes the conclusion of this study.

2. Infectious Disease Principles and Characteristics of Health Food Safety Scare Behavior Diffusion

2.1. Adaptability Analysis of the Epidemic Model. The epidemic model, as a classical model of virus transmission, has been extensively used in studying social behavior diffusion [3–5, 39–41]. The source of an infectious disease is a virus carrier or a pathogen, which spreads its own virus by a contact through a certain medium [42]. Health food safety scare behavior diffusion, what we define is the interaction and transmission of the behavior of health food consumers. It denotes that the scare behavior of health food consumers will spread through all kinds of media to health food consumers who are uninfected. The health food safety scare behavior affects stakeholders like a virus, and many mechanisms are similar between the processes of diffusion and virus propagation. The principal representations are as follows:

(1) Pathogen-diffusion source. The spread of health food safety scare mainly stems from the concerns of health food consumers over health food safety [43]. With the influence of factors, such as health food information transparency, health food consumers' ability to process information, and network structure, health food consumers who produce health food safety scare behavior are "pathogens" or diffusion sources that have the potential to spread. A diffusion source, which is a prerequisite for health food safety scare behavior diffusion, will spread health food safety scare behavior to health food consumers through the diffusion media, thereby presenting a significant herding effect.

(2) Contagion medium-diffusion medium. A diffusion medium is the carrier of the diffusion source, such as the Internet, mobile phone, TVs, and other mass media, and face-to-face communication between health food consumers. The health food safety information transmitted through the diffusion media is related to the health and safety of health food consumers. The transparency of health food safety information affects the confidence of health food consumers in health food safety [3–5, 16, 17].

(3) Infectiousness. Health food consumers involved in the health food safety scare deliver their psychological cognition, behavior deviation, and other information to health food
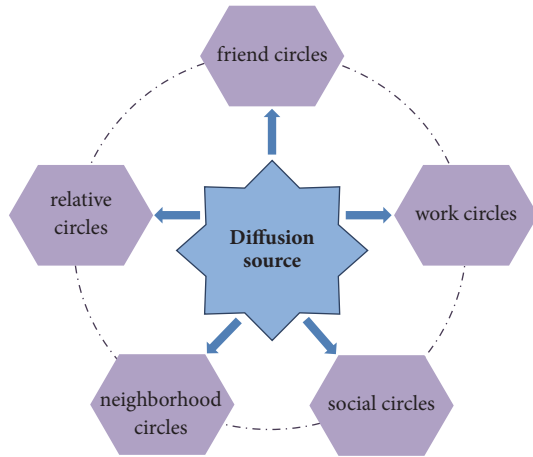
Figure 2: Diffusion media and diffusion path of health food safety scare behavior.

Figure 3: Health food safety scare behavior diffusion mechanism.

consumers who are in a healthy state through the diffusion media given the influence of health food safety information transparency, thereby inducing the latter to divert their attention and psychological cognition to the health food safety accident and consequently causing scare behavior. This mechanism shows that the health food safety scare behavior has certain infectiousness. In Figure 2, health food consumers that exhibit health food safety scare behavior will pass information, such as their own psychological state and behavioral biases to the external environment, thus affecting the health food consumers in a normal state and causing them to panic.

(4) Immunity. In the epidemic model, individuals are immune to pathogens. Health food consumers show different immunity levels to health food safety scare behavior given their different degrees of psychological quality and health food safety knowledge as influenced by the health food safety regulatory information transparency. If the health food information transparency is considerable, the psychological quality of health food consumers is high, and the health food safety knowledge of health food consumers abounds, then the health food safety scare behavior diffusion will be suppressed. Otherwise, the health food safety scare behavior diffusion will be accelerated.

Therefore, the scare behavior diffusion process of health food industry emergencies is characterized by the infectious disease propagation process, which can be used to analyze and simulate the process of the scare behavior diffusion for health food. Therefore, constructing an SIRS epidemic model of the health food safety scare behavior diffusion is reasonable.

2.2. Health Food Safety Scare Behavior Diffusion Mechanism. In a complex network of N health food consumers, the nodes represent health food consumers, and the edges represent the diffusion media between two health food consumers. Several health food consumers exhibit scare behavior, whereas other health food consumers have the potential of adopting the behavior. Accordingly, spreading the health food safety scare
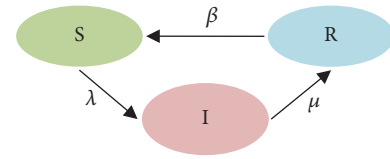
behavior is realized through such a complex network. Degree k represents the relationship between health food consumers and also denotes the number of diffusion media and paths.

In the network of health food consumers, each health food consumer exhibits one of the following states:

(1) Normal state S. This state indicates that health food consumers are unaffected by the health food safety scare behavior but may contract it.

(2) Scare state I. This state indicates that health food consumers have contracted the health food safety scare behavior by diffusion and are contagious.

(3) Temporary immunity state R. This state indicates that health food consumers have recovered to normal states and have the ability to be immune from health food safety scare behavior temporarily; however, these consumers will be in a vulnerable normal state again after a time and continue to be affected by the health food safety scare behavior.

At the beginning, no health food safety incident has occurred in the health food consumer network. A health food consumer in the network is either in the normal state S, scare state I, or temporary immunity state R when a health food emergency occurs. At moment t, the proportion of health food consumers who are in the normal state in the network is called the normal state health food consumer density, which is denoted by $s(t)$. The proportion of the health food consumers who are in the scare state in the network is called the scare state health food consumer density, which is denoted by $i(t)$. The proportion of health food consumers who are in the temporary immunity state in the network is called the temporary immunity state health food consumer density, which is denoted by $r(t)$. Therefore, $s(t) + i(t) + r(t) = 1$. The network of health food consumers tends to be balanced when $t \longrightarrow \infty$. I is the origin of diffusion that spreads to S at probability λ (diffusion rate) and recovers to be R at probability μ (recovery rate) simultaneously. R loses immunity at probability β (immune failure rate) to be S (Figure 3).

The health food safety scare behavior diffusion between health food consumers is uneven. This unevenness is related to the behavior diffusion rate, the health food safety information transparency, and the health food consumers' ability to process information.

(1) Health food safety information transparency. According to Chen et al. [3–5], the transparency of food safety information is an important factor that affects food safety risk. Thus, for health food, a high transparency of health food safety information denotes an improved delivery or disclosure of the information on health food by the government, enterprises, and media. The spread of health food

Table 1: Sensitivity analysis of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k to the health food safety scare behavior diffusion rate λ_k.

θ	η_k									Expectation	Variance
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
0.1	5.60E-01	3.39E-01	2.10E-01	1.33E-01	8.48E-02	5.46E-02	3.54E-02	2.31E-02	1.52E-02	1.62E-01	3.34E-02
0.2	4.82E-01	2.81E-01	1.72E-01	1.09E-01	7.10E-02	4.69E-02	3.14E-02	2.12E-02	1.45E-02	1.37E-01	2.42E-02
0.3	3.99E-01	2.25E-01	1.38E-01	8.85E-02	5.87E-02	3.99E-02	2.77E-02	1.95E-02	1.39E-02	1.12E-01	1.62E-02
0.4	3.15E-01	1.73E-01	1.07E-01	7.01E-02	4.79E-02	3.37E-02	2.43E-02	1.78E-02	1.33E-02	8.91E-02	9.79E-03
0.5	2.33E-01	1.28E-01	8.03E-02	5.43E-02	3.85E-02	2.82E-02	2.12E-02	1.63E-02	1.27E-02	6.80E-02	5.19E-03
0.6	1.60E-01	8.90E-02	5.81E-02	4.11E-02	3.05E-02	2.34E-02	1.84E-02	1.48E-02	1.21E-02	4.97E-02	2.31E-03
0.7	9.95E-02	5.83E-02	4.04E-02	3.02E-02	2.37E-02	1.92E-02	1.60E-02	1.35E-02	1.15E-02	3.47E-02	8.13E-04
0.8	5.47E-02	3.55E-02	2.68E-02	2.16E-02	1.82E-02	1.56E-02	1.37E-02	1.22E-02	1.10E-02	2.33E-02	2.00E-04
0.9	2.58E-02	1.98E-02	1.69E-02	1.50E-02	1.36E-02	1.26E-02	1.18E-02	1.11E-02	1.05E-02	1.52E-02	2.46E-05

safety scare behavior will be suppressed when health food consumers provide sufficient information. Otherwise, the spread of health food safety scare behavior will be promoted.

(2) Health food consumers' ability to process information. On the one hand, health food consumers exhibit different cognitive performances in terms of health food. Several health food consumers are unaffected by the scare behavior and inhibit the spread of this behavior given their knowledge of health food. However, several health food consumers are vulnerable to the influence of connected health food consumers, thereby demonstrating a remarkable herd behavior. On the other hand, different health food consumers possess various mental qualities and therefore show different psychological qualities and psychological cognitions. Health food consumers with high psychological quality are less or even unaffected by scare behavior. Therefore, a high level of health food consumers' ability to process information exerts a certain inhibitory effect on the health food safety scare behavior diffusion.

The given analysis shows that the diffusion rate in the proposed model of the health food safety scare behavior is as follows:

$$\lambda_k(\lambda_0, \theta, \eta) = \lambda_0^{\theta^{1-\eta_k}}, \tag{1}$$

where λ_0 denotes the diffusion rate that satisfies the condition $0 < \lambda_0 < 1$ and θ denotes the health food safety information transparency that satisfies the condition $0 < \theta < 1$. A high θ indicates a transparent health food safety information. η_k denotes the degree k of health food consumers' ability to process information. A high η_k denotes that health food consumers can process information. To better analyze the diffusion rate mode, we make visualization of the model via Matlab2016b. Based on that, we also make the sensitivity analysis. The results are shown as in Figure 4 and Table 1 .

Figure 4 depicts the effects of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k on the health food safety scare behavior diffusion rate λ_k. The health food safety scare behavior diffusion rate λ_k decreases nonlinearly with an increase in the health food safety information transparency θ or the degree k of health food consumers' ability to

Figure 4: Effects of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k on the health food safety scare behavior diffusion rate λ_k when $\lambda_0 = 0.01$ and $k = 1000$.

process information η_k. In the sensitivity analysis displayed in Table 1, the health food safety information transparency θ reflects a marginal diminishing decreasing characteristic in relation to the health food safety scare behavior diffusion rate λ_k, whereas the degree k of health food consumers' ability to process information η_k shows a marginal incremental decreasing characteristic in relation to the health food safety scare behavior diffusion rate λ_k. These results are consistent with the facts. Therefore, the established diffusion rate model on health food safety scare can be reasonably constructed.

3. Diffusion Model of Health Food Safety Scare Behavior

3.1. Model Construction. If the density of degree k of health food consumers in the scare state is $i_k(t)$ at time t, then the density of its equilibrium state is $i_k(\infty)$. $s_k(\infty)$ and $r_k(\infty)$ have the same definition. According to mean-field theory (Moreno et al., 2003; Yang et al., 2006), the dynamics model

of health food safety scare behavior diffusion in this network is as follows:

$$s_k'(t) = -k\lambda_k \Theta(t) s_k(t) + \beta r_k(t)$$

$$i_k'(t) = k\lambda_k s_k(t) \Theta(t) - \mu i_k(t) \tag{2}$$

$$r_k'(t) = \mu i_k(t) - \beta r_k(t),$$

where $\Theta(t)$ represents the probability that a health food consumer in the normal state is directly linked to a health food consumer in the scare state. In equilibrium, the stable value is denoted by $\Theta(\infty)$. The first term on the right side of the first line of (2) is the health food safety scare behavior production item. This term is proportional to the degree k of health food consumers, the health food safety scare behavior diffusion rate λ_k, and the density $s_k(t)$ of degree k of health food consumers in the normal state at time t. The second term is the health food safety scare behavior annihilation term, which is proportional to the immune failure rate β and the density $r_k(t)$ of degree k of health food consumers in the temporary immunity state at time t. $\Theta(t)$ is defined as follows:

$$\Theta(t) = \frac{\sum_k k P(k) i_k(t)}{\langle k \rangle}. \tag{3}$$

Equation (3) represents the probability that a health food consumer in the normal state is directly linked to a health food consumer in the scare state at time t; $\langle k \rangle = \sum_k k P(k)$ denotes the average degree of the network.

Setting $\begin{cases} s_k'(t)=0 \\ i_k'(t)=0 \\ r_k'(t)=0 \end{cases}$ yields the following unsteady solution:

$$i_k(\infty) = \frac{\lambda_k k \Theta(\infty)}{\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu}. \tag{4}$$

The following equation can be obtained by substituting (4) into (3):

$$\Theta(\infty) = \frac{\sum_k k P(k)}{\langle k \rangle} \frac{\lambda_k k \Theta(\infty)}{\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu} \tag{5}$$

$$= G(\Theta(\infty)),$$

where $\Theta(\infty) = 0$ is a trivial solution in (5); that is, the health food safety scare behavior does not spread widely. If the health food safety scare behavior diffusion occurs, then (5) must have a nontrivial solution that satisfies $(dG(\Theta(\infty))/d\Theta(\infty))|_{\Theta(\infty)=0} \geq 1$. $\lambda_k \langle k^2 \rangle / \mu \langle k \rangle \geq 1$ can be obtained through simplification. Therefore, the critical condition of the health food safety scare behavior diffusion rate can be obtained as follows:

$$\lambda_0^{\theta^{1-\eta_k}} = \frac{\mu \langle k \rangle}{\langle k^2 \rangle}. \tag{6}$$

Accordingly, we can obtain the threshold of the rate of health food safety scare behavior diffusion $\lambda_0* = (\mu \langle k \rangle / \langle k^2 \rangle)^{\theta^{\eta_k-1}}$, where $\langle k^2 \rangle = \sum_k k^2 P(k)$. Finally, we can derive the entire health food consumer network diffusion as follows:

$$i(\infty) = \sum_k P(k) i_k(\infty). \tag{7}$$

3.2. Model Analysis.
In accordance with the established model in Section 3.1, this study analyzes the threshold of the rate of health food safety scare behavior diffusion; that is, $\lambda_0* = (\mu \langle k \rangle / \langle k^2 \rangle)^{\theta^{\eta_k-1}}$, in accordance with the following relationships.

(1) Relationship to the Health Food Safety Information Transparency θ

$$\frac{\partial \lambda_0*}{\partial \theta} = (\eta_k - 1) \cdot \left(\frac{\mu \langle k \rangle}{\langle k^2 \rangle} \right)^{\theta^{\eta_k-1}} \cdot \ln\left(\frac{\mu \langle k \rangle}{\langle k^2 \rangle} \right) \cdot \theta^{\eta_k-2} \tag{8}$$

$$> 0$$

(2) Relationship to the Degree k of Health Food Consumers' Ability to Process Information η_k

$$\frac{\partial \lambda_0*}{\partial \eta_k} = \left(\frac{\mu \langle k \rangle}{\langle k^2 \rangle} \right)^{\theta^{\eta_k-1}} \cdot \ln\left(\frac{\mu \langle k \rangle}{\langle k^2 \rangle} \right) \cdot \theta^{\eta_k-1} \ln \theta > 0 \tag{9}$$

(3) Relationship to the Recovery Rate μ of Health Food Consumers

$$\frac{\partial \lambda_0*}{\partial \mu} = \left(\frac{\langle k \rangle}{\langle k^2 \rangle} \right)^{\theta^{\eta_k-1}} \cdot \theta^{\eta_k-1} \cdot \mu^{\theta^{\eta_k-1}-1} > 0 \tag{10}$$

$$\frac{\partial^2 \lambda_0*}{\partial \mu^2} = \left(\frac{\langle k \rangle}{\langle k^2 \rangle} \right)^{\theta^{\eta_k-1}} \cdot \theta^{\eta_k-1} \cdot \left(\theta^{\eta_k-1} - 1 \right) \mu^{\theta^{\eta_k-1}-2} \tag{11}$$

$$> 0$$

In summary, the following theorem is obtained.

Theorem 1. *The threshold λ_0* of the rate of health food safety scare behavior diffusion is the monotonically increas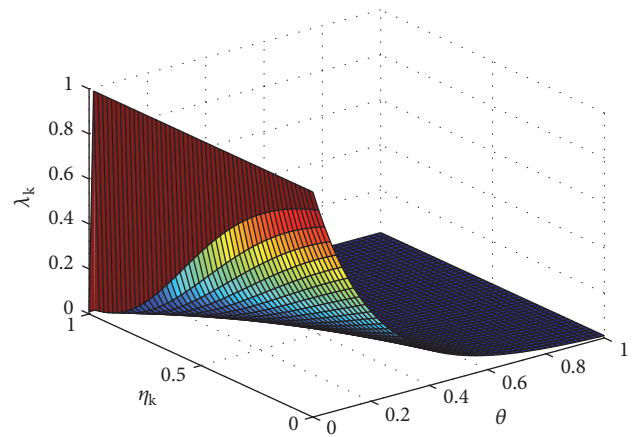ing function of the health food safety information transparency θ, the degree k of health food consumers' ability to process information η_k, and the monotonically increasing convex function of the recovery rate μ of health food consumers.*

(4) Relationship to Different Network Structures. Let P' and P represent the degree distribution of the two health food consumer networks K' and K, respectively.

$G'(\Theta(\infty)) = \langle k^2 \rangle \lambda_k \mu / \langle k \rangle (\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu)^2 > 0$ and $G''(\Theta(\infty)) = -\langle k^2 \rangle \lambda_k^2 k(1 + \mu/\beta)/\langle k \rangle (\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu)^4 < 0$. Thus, $G(\Theta(\infty))$ is the monotonically increasing concave function of $\Theta(\infty)$. $G(1) = (\sum_k k P(k)/\langle k \rangle)(\lambda_k k/(\lambda_k k(1+\mu/\beta)+\mu)) < (\sum_k k P(k)/\langle k \rangle)(\lambda_k k/\lambda_k k) = 1$ and $G(0) = 0$. Thus, $G(\Theta(\infty))$ has at least one fixed point in the interval $[0, 1]$. Health food safety scare behavior spreads to the entire consumer network, and the network reaches the equilibrium state when $\lambda_0 > \lambda_0*$. Simultaneously, $G'(\Theta(\infty))|_{\Theta=0} > 1$. Therefore, $\Theta(\infty) = (\langle k^2 \rangle/\langle k \rangle)(\lambda_k \Theta(\infty)/(\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu))$ has the only equilibrium point in the interval $[0, 1]$, and $\Theta(\infty)^* > 0$.

The following theorems can be obtained in accordance with the criteria of stochastic dominance.

Theorem 2. *If the average degree $\langle K' \rangle$ of the health food consumer network K' is greater than the average degree $\langle K \rangle$ of*

the health food consumer network K, then (1) the equilibrium value $\Theta(\infty)^{*\prime}$ of the health food safety scare behavior diffusion in network K' is greater than the equilibrium value $\Theta(\infty)^*$ of the health food safety scare behavior diffusion in network K when $\lambda_0 > \lambda_0*$; moreover, (2) the scale $i(\infty)^{*\prime}$ of the health food safety scare behavior diffusion in network K' is greater than the scale $i(\infty)^*$ of the health food safety scare behavior diffusion in network K.

Proof. Theorem 2(1) is untenable. In particular, if the average degree $\langle K' \rangle$ of the health food consumer network K' is greater than the average degree $\langle K \rangle$ of the health food consumer network K, then $\Theta(\infty)^{*\prime} \leq \Theta(\infty)^*$; that is, $\Theta(\infty)^* \geq \Theta(\infty)^{*\prime} = G_k'(\Theta(\infty)^*)$. The equilibrium value $\Theta(\infty)^*$ of the health food safety scare behavior diffusion is unique, and $\Theta(\infty)^* > 0$ when $\lambda_0 > \lambda_0*$. If $H(k) = \lambda_k k \Theta(\infty)/\langle k \rangle [\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu]$, then we can obtain $\partial H(k)/\partial k = \mu \lambda_k \Theta(\infty)/(\langle k \rangle [\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu])^2 > 0$. Thus, $H(k)$ is the monotonically increasing function of k. According to Jackson [37], if the average degree $\langle K' \rangle$ of network K' is greater than the average degree $\langle K \rangle$ of network K, then P' first-order stochastic dominates P. Thus, if the average degree $\langle K' \rangle$ of the health food consumer network K' is greater than the average degree $\langle K \rangle$ of the health food consumer network K, then $\sum_k H(k)P'(k) > \sum_k H(k)P(k)$. According to (5), $\forall \Theta(\infty) > 0$, $G_k'(\Theta(\infty)) > G_k(\Theta(\infty))$ can be obtained. Accordingly, $\forall \Theta(\infty)^* > 0$, $G_k'(\Theta(\infty)^*) > G_k(\Theta(\infty)^*)$ and $\Theta(\infty)^* \geq G_k'(\Theta(\infty)^*) > G_k(\Theta(\infty)^*)$ can be obtained. This result contradicts $\Theta(\infty)^* = G_k(\Theta(\infty)^*)$. Thus, our hypothesis is tenable, and Theorem 2(1) is true.

According to Theorem 2(1), $\Theta(\infty)^{*\prime} > \Theta(\infty)^*$ when the average degree $\langle K' \rangle$ of the health food consumer network K' is greater than the average degree $\langle K \rangle$ of the health food consumer network K. Clearly, $i_k(\infty) = \lambda_k k \Theta(\infty)/(\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu)$ is the monotonically increasing concave function of $\Theta(\infty)$. Thus, for any $k > 0$, $i_k(\infty)^{*\prime} > i_k(\infty)^*$ and $\sum_k i_k(\infty)^{*\prime} > \sum_k i_k(\infty)^*$ can be obtained. Furthermore, $\sum_k P'(k)i_k(\infty)^{*\prime} > \sum_k P'(k)i_k(\infty)^*$, that is, $i(\infty)^{*\prime} > \sum_k P'(k)i_k(\infty)^*$. Apparently, $i_k(\infty) = \lambda_k k \Theta(\infty)/(\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu)$ is the monotonically increasing concave function of k. Jackson [37] noted that if the average degree $\langle K' \rangle$ of network K' is greater than the average degree $\langle K \rangle$ of network K, then P' first-order stochastic dominates P. Therefore, $\sum_k P'(k)i_k(\infty)^* > \sum_k P(k)i_k(\infty)^*$ and $i(\infty)^{*\prime} > \sum_k P'(k)i_k(\infty)^* > \sum_k P(k)i_k(\infty)^* = i(\infty)^*$ can be obtained. Thus, Theorem 2(2) is proven. □

Theorem 3. *If the heterogeneity of the health food consumer network K' is greater than the heterogeneity of the health food consumer network K, then the equilibrium value $\Theta(\infty)^{*\prime}$ of the health food safety scare behavior diffusion in network K' is lower than the equilibrium value $\Theta(\infty)^*$ of the health food safety scare behavior diffusion in network K when $\lambda_0 > \lambda_0*$.*

Proof. Theorem 3 is untenable. In particular, if the heterogeneity of the health food consumer K' is higher than the heterogeneity of the health food consumer network K, then $\Theta(\infty)^* \leq \Theta(\infty)^{*\prime} = G_k'(\Theta(\infty)^*)$ can be obtained. Furthermore, $\Theta(\infty)$ has the only equilibrium point in the interval

$[0, 1]$ and $\Theta(\infty)^* > 0$. $\partial^2 H(k)/\partial k^2 = -2\langle k \rangle \lambda_k \Theta(\infty)(1 + \mu/\beta)/(\langle k \rangle [\lambda_k k \Theta(\infty)(1 + \mu/\beta) + \mu])^3 < 0$ is yielded. Thus, $H(k)$ is a concave function of k. According to Jackson [37], if the heterogeneity of the network K' is greater than the heterogeneity of the network K, then P second-order stochastic dominates P'. Therefore, $\sum_k H(k)P(k) > \sum_k H(k)P'(k)$ can be obtained. According to (5), $\forall \Theta(\infty) > 0$, $G_k(\Theta(\infty)) > G_k'(\Theta(\infty))$. Thus, $\forall \Theta(\infty)^* > 0$, $G_k(\Theta(\infty)^*) > G_k'(\Theta(\infty)^*)$. Accordingly, $\Theta(\infty)^* \leq \Theta(\infty)^{*\prime} = G_k'(\Theta(\infty)^*) < G_k(\Theta(\infty)^*)$. This finding contradicts $\Theta(\infty)^* = G_k(\Theta(\infty)^*)$. Thus, our hypothesis is tenable, and Theorem 3 is true. □

4. Analogue Simulation

Numerical simulation analysis is the most effective means of testing real-time dynamic data without the requirement for numerous empirical validations. Therefore, following Chen et al. [3–5] and He et al. [44], we simulate the health food safety scare behavior diffusion by using MATLAB 2016b software.

As the network of health food consumers is a complex network, it is very difficult to obtain the characteristics or the real datasets of the network. Therefore, we selected the most representative three heterogeneous networks to theoretically study the health food safety scare behavior diffusion. BA scale-free network [20], WS network [19], and Exponential network were used. We used BA network, WS network, and Exponential network to describe the feature of different network structure. In fact, fewer nodes have many direct connections with other nodes in BA network, but a large number of nodes have various direct connections with other nodes in WS network. And the heterogeneity of Exponential network is between them.

In the three network models, the degree distribution of BA network is $P(k) \propto 2m^2/k^3$, the degree distribution of long distance connection of nodes in WS network is equal to 0.05, and the degree distribution of Exponential network is $P(k) \propto e^{-\varepsilon k/2m}$. Then, let $N = 1000$, and let $m_0 = m = 5$, $\varepsilon = 2$.

4.1. Analysis of the Threshold of the Rate of Health Food Safety Scare Behavior Diffusion. We analyze the threshold of the rate of health food safety scare behavior diffusion to depict the evolution characteristics of the health food safety scare behavior diffusion under the influence of the health food safety information transparency and health food consumers' ability to process information.

Figure 5 demonstrates the effects of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k on the threshold λ_0* of the rate of health food safety scare behavior diffusion under three network structures. Theorem 1 is verified visually in Figure 5; that is, the threshold λ_0* of the rate of health food safety scare behavior diffusion is the monotonically increasing function of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k. An increase in the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k increases the threshold λ_0* of the rate of health food

(a) BA network

(b) WS network

(c) Exponential network
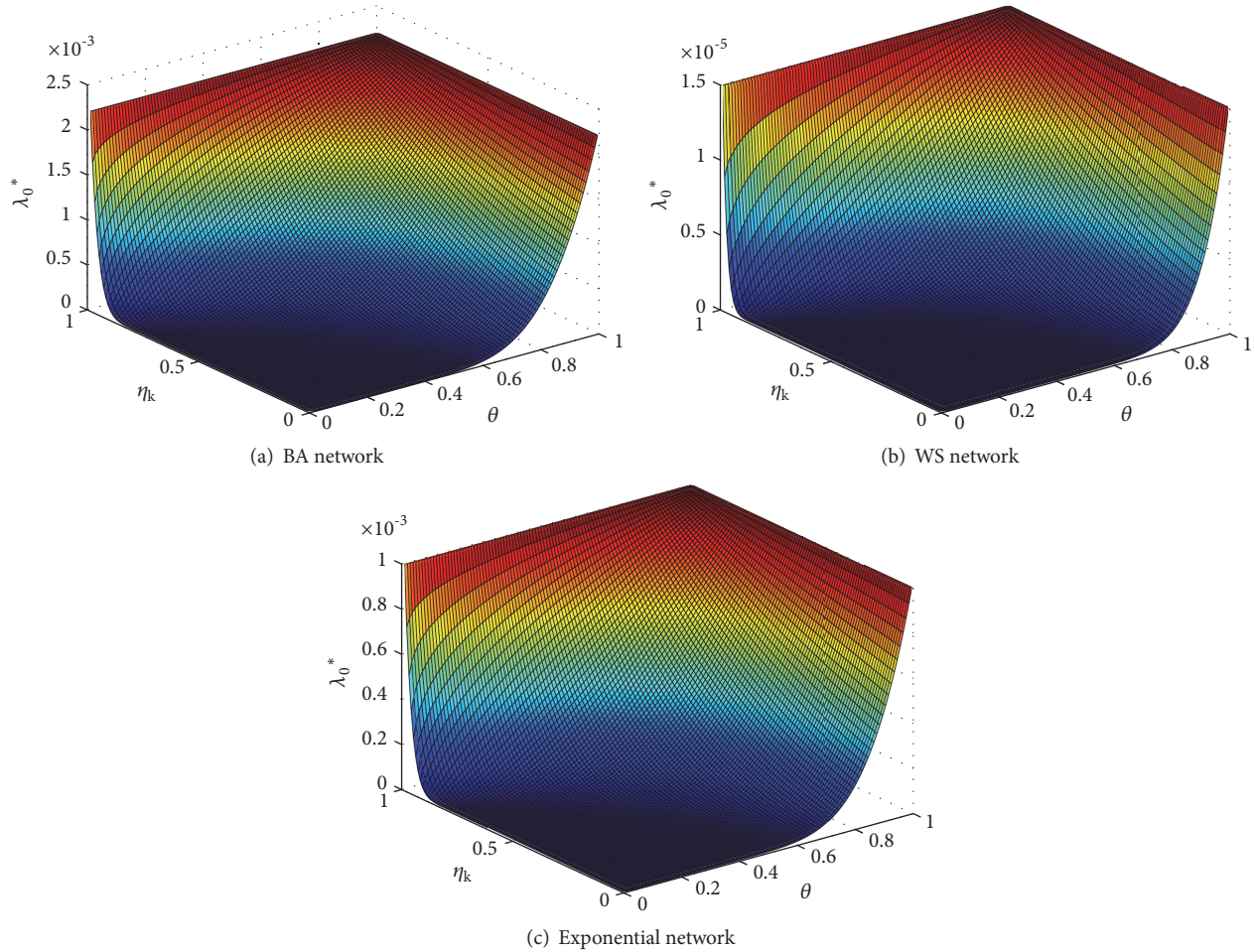
FIGURE 5: Change rules in the threshold of the rate of health food safety scare behavior diffusion under three network structures: (a) BA network, which exerts the effects of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k on the threshold of the rate of health food safety scare behavior diffusion λ_0*; (b) WS network, which shows the effects of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k on the threshold of the rate of health food safety scare behavior diffusion λ_0*; (c) Exponential network, which shows the effects of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k on the threshold of the rate of health food safety scare behavior diffusion λ_0* (where $k = 1000$ and $\mu = 0.01$).

safety scare behavior diffusion because a high health food safety information transparency indicates comprehensive information on health food in the market for health food consumers. In this scenario, spreading scare behavior when a health food safety incident occurs is difficult. Similarly, a high health food consumers' ability to process information denotes a rational analysis of the health food safety scare behavior. Consequently, no herd behavior will emerge, and health food consumers will inhibit spreading or refuse to spread scare behavior.

A comparison of Figures 5(a), 5(b), and 5(c) indicates that the thresholds λ_0* of the rate of health food safety scare behavior diffusion differ between the three network structures. The threshold of the rate of health food safety scare behavior diffusion in the WS network is much lower than the threshold of the rate of health food safety scare behavior diffusion in the BA network and the Exponential

network; the threshold of the rate of health food safety scare behavior diffusion in the Exponential network is lower than the threshold of the rate of health food safety scare behavior diffusion in the BA network. These results suggest that the health food safety scare behavior in the BA network is more difficult to spread. This phenomenon is attributed to the improved heterogeneity of the BA network. In a network with high heterogeneity, the heterogeneity of the connecting edges of health food consumers and the resistance to and spread of information are also high. Thus, spreading health food safety scare behavior is difficult. This phenomenon also reveals that the structural characteristic of a network significantly affects the health food safety scare behavior diffusion.

To better analyze the threshold of the rate of health food safety scare behavior diffusion to depict the evolution characteristics of the health food safety scare behavior diffusion under the influence of the health food safety information

TABLE 2: Sensitivity analysis of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k to the rate of health food safety scare behavior diffusion λ_0* in the BA network.

θ	η_k									Expectation	Variance
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
0.1	4.79E-39	2.92E-21	5.54E-15	9.90E-12	9.95E-10	2.29E-08	2.24E-07	1.27E-06	4.97E-06	7.20E-07	2.71E-12
0.2	3.65E-31	3.30E-18	2.29E-13	9.10E-11	3.99E-09	5.50E-08	3.83E-07	1.71E-06	5.64E-06	3.52E-12	3.52E-12
0.3	6.64E-25	1.31E-15	6.23E-12	6.89E-10	1.46E-08	1.27E-07	6.42E-07	2.29E-06	6.41E-06	4.59E-12	4.59E-12
0.4	6.23E-20	2.13E-13	1.16E-10	4.37E-09	4.88E-08	2.79E-07	1.06E-06	3.05E-06	7.26E-06	6.00E-12	6.00E-12
0.5	5.56E-16	1.63E-11	1.56E-09	2.36E-08	1.51E-07	5.92E-07	1.71E-06	4.04E-06	8.22E-06	7.88E-12	7.88E-12
0.6	7.63E-13	6.55E-10	1.55E-08	1.10E-07	4.31E-07	1.21E-06	2.73E-06	5.31E-06	9.29E-06	1.04E-11	1.04E-11
0.7	2.37E-10	1.52E-08	1.19E-07	4.46E-07	1.15E-06	2.38E-06	4.28E-06	6.95E-06	1.05E-05	1.37E-11	1.37E-11
0.8	2.26E-08	2.21E-07	7.29E-07	1.61E-06	2.87E-06	4.54E-06	6.59E-06	9.03E-06	1.18E-05	1.77E-11	1.77E-11
0.9	8.45E-07	2.16E-06	3.62E-06	5.16E-06	6.76E-06	8.38E-06	1.00E-05	1.17E-05	1.33E-05	1.86E-11	1.86E-11

TABLE 3: Sensitivity analysis of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k to the rate of health food safety scare behavior diffusion λ_0* in the WS network.

θ	η_k									Expectation	Variance
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
0.1	7.66E-22	4.83E-12	1.39E-08	8.63E-07	1.09E-05	6.17E-05	2.16E-04	5.63E-04	1.19E-03	2.28E-04	1.66E-07
0.2	1.69E-17	2.33E-10	1.08E-07	2.93E-06	2.35E-05	9.99E-05	2.91E-04	6.64E-04	1.28E-03	2.63E-04	1.95E-07
0.3	4.75E-14	6.29E-09	6.69E-07	8.94E-06	4.80E-05	1.58E-04	3.87E-04	7.80E-04	1.37E-03	3.06E-04	2.29E-07
0.4	2.61E-11	1.04E-07	3.35E-06	2.47E-05	9.35E-05	2.44E-04	5.10E-04	9.14E-04	1.47E-03	3.63E-04	2.69E-07
0.5	3.92E-09	1.14E-06	1.40E-05	6.26E-05	1.74E-04	3.70E-04	6.65E-04	1.07E-03	1.58E-03	4.37E-04	3.15E-07
0.6	2.10E-07	8.70E-06	4.98E-05	1.46E-04	3.11E-04	5.48E-04	8.59E-04	1.24E-03	1.69E-03	5.39E-04	3.66E-07
0.7	4.97E-06	4.92E-05	1.53E-04	3.17E-04	5.34E-04	7.97E-04	1.10E-03	1.44E-03	1.80E-03	6.88E-04	4.14E-07
0.8	6.12E-05	2.15E-04	4.15E-04	6.41E-04	8.84E-04	1.14E-03	1.40E-03	1.66E-03	1.93E-03	9.26E-04	4.26E-07
0.9	4.50E-04	7.54E-04	1.00E-03	1.22E-03	1.42E-03	1.59E-03	1.76E-03	1.91E-03	2.06E-03	1.35E-03	2.94E-07

TABLE 4: Sensitivity analysis of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k to the rate of health food safety scare behavior diffusion λ_0* in the Exponential network.

θ	η_k									Expectation	Variance
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
0.1	3.95E-22	9.16E-13	3.08E-09	2.31E-07	3.43E-06	2.19E-05	8.49E-05	2.40E-04	5.44E-04	9.94E-05	3.40E-08
0.2	6.85E-18	5.03E-11	2.69E-08	8.50E-07	7.78E-06	3.67E-05	1.16E-04	2.85E-04	5.84E-04	1.15E-04	3.99E-08
0.3	1.72E-14	1.55E-09	1.84E-07	2.80E-06	1.67E-05	6.01E-05	1.58E-04	3.38E-04	6.26E-04	1.34E-04	4.68E-08
0.4	9.15E-12	2.90E-08	1.02E-06	8.33E-06	3.42E-05	9.60E-05	2.12E-04	3.99E-04	6.71E-04	1.58E-04	5.51E-08
0.5	1.41E-09	3.56E-07	4.70E-06	2.26E-05	6.69E-05	1.50E-04	2.82E-04	4.70E-04	7.19E-04	1.91E-04	6.47E-08
0.6	7.96E-08	3.04E-06	1.82E-05	5.63E-05	1.25E-04	2.29E-04	3.71E-04	5.51E-04	7.69E-04	2.36E-04	7.55E-08
0.7	2.03E-06	1.91E-05	6.09E-05	1.30E-04	2.25E-04	3.44E-04	4.84E-04	6.45E-04	8.23E-04	3.04E-04	8.60E-08
0.8	2.72E-05	9.17E-05	1.78E-04	2.78E-04	3.88E-04	5.05E-04	6.26E-04	7.52E-04	8.80E-04	4.14E-04	8.90E-08
0.9	2.18E-04	3.52E-04	4.62E-04	5.59E-04	6.47E-04	7.28E-04	8.03E-04	8.73E-04	9.40E-04	6.20E-04	5.93E-08

transparency and health food consumers' ability to process information, we make the sensitivity analysis within the three networks. The results are shown as in Tables 2, 3, and 4.

The sensitivity analysis results in Tables 2, 3, and 4 show that the threshold λ_0* of the rate of health food safety scare behavior diffusion is the monotonically increasing function of the health food safety information transparency θ and the degree k of health food consumers' ability to process information η_k. Furthermore, the health food safety information transparency θ shows the characteristics of incremental margins in relation to the threshold λ_0* of the rate of health food safety scare behavior diffusion, whereas the degree k of health food consumers' ability to process information η_k denotes the characteristics of diminishing margins in relation to the threshold λ_0* of the rate of health food safety scare behavior diffusion.
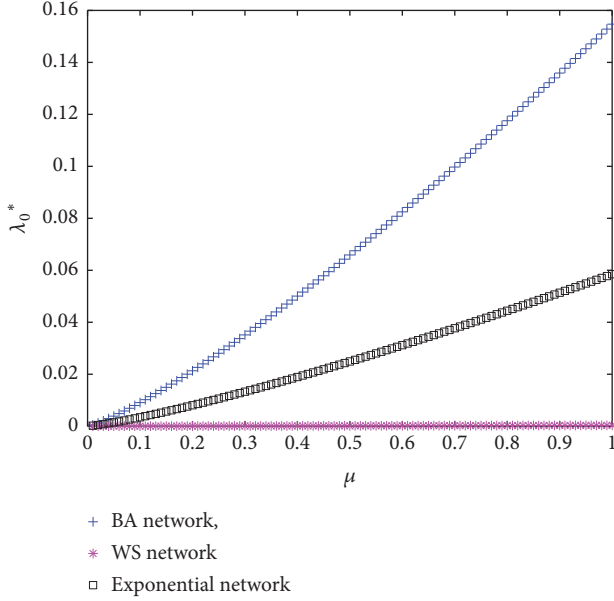
FIGURE 6: Effects of the recovery rate μ of health food consumers on the threshold λ_0* of the rate of health food safety scare behavior diffusion under three network structures (where $\theta = 0.5$, $\eta_k = 0.7$, and $k = 1000$).

Then, we simulate the relationship between the recovery rate μ and the threshold λ_0* of the rate of health food safety scare behavior diffusion. The result is shown as in Figure 6.

Figure 6 displays the effects of the recovery rate μ of health food consumers on the threshold λ_0* of the rate of health food safety scare behavior diffusion under three network structures. Figure 6 verifies the conclusion of Theorem 1, that is, the threshold λ_0* of the rate of health food safety scare behavior diffusion is the monotonically increasing function of the recovery rate μ of health food consumers. In particular, the threshold λ_0* of the rate of health food safety scare behavior diffusion increases with the recovery rate μ of health food consumers, and the rate is low. This outcome is due to a high recovery rate of health food consumers indicates their improved ability to resist scare behavior. Thus, the threshold required to achieve the health food safety scare behavior diffusion must be high. In addition, Figure 6 illustrates that the threshold of the rate of health food safety scare behavior diffusion is much lower in the WS network than in the BA network and Exponential network. Furthermore, the former tends to zero. And the change trend in the Exponential network is much smaller in the BA network, while which in the WS network is the smallest.

4.2. Analysis of the Scale of the Health Food Safety Scare Behavior Diffusion. Subsequently, we analyze the scale of the health food safety scare behavior diffusion. The results are shown as in Figure 7.

Figure 7 presents the effects of the network structure of health food consumers, the health food safety information transparency θ, the degree k of health food consumers' ability to process information η_k, and the degree k of health food

consumers on the scale $i(\infty)$ of the health food safety scare behavior diffusion. Figures 7(a)–7(f) demonstrate that a high heterogeneity of a health food consumer network denotes a low scale of health food safety scare behavior diffusion.

In Figures 7(a) and 7(c), with increasing in the health food safety information transparency θ, the scale $i(\infty)$ of the health food safety scare behavior diffusion decreases. That is, health food safety information transparency exerts inhibitory effects on health food safety scare behavior diffusion and shows the characteristics of incremental margins. In addition, the inhibitory effect of health food safety information transparency on health food safety scare behavior diffusion is apparent in the BA network because health food consumers exhibit few direct edges in the BA network, and a low health food safety information transparency can exert a strong inhibitory effect on the health food safety scare behavior diffusion. By contrast, in the WS network, the direct edges of health food consumers are relatively large, and their similarity is high; moreover, a low health food safety information transparency can hardly exert a strong inhibitory effect on the health food safety scare behavior diffusion. As for the Exponential network, the situation is between the BA network and WS network.

In Figures 7(b) and 7(d), the scale $i(\infty)$ of the health food safety scare behavior diffusion decreases with an increase in the degree k of health food consumers' ability to process information η_k. That is, the degree k of health food consumers' ability to process information η_k exerts an inhibitory effect on the health food safety scare behavior diffusion and shows the characteristics of incremental margins. Similarly, a high health food consumers' ability to process information denotes a strong inhibitory effect on the health food safety scare behavior diffusion.

In Figures 7(e) and 7(f), the degree k of health food consumers indicates a marginal diminishing rising characteristic in relation to the scale $i(\infty)$ of the health food safety scare behavior diffusion. Moreover, the scales of the health food safety scare behavior diffusion under different networks tend to a steady value with an increase in the degree k of health food consumers.

Last, we simulate the relationship between the recovery rate μ and the scale $i(\infty)$ of the health food safety scare behavior diffusion. The result is shown as in Figure 8.

Figure 8 illustrates the effects of the recovery rate μ of health food consumers on the scale $i(\infty)$ of the health food safety scare behavior diffusion under three kinds of network structures. The scale $i(\infty)$ of the health food safety scare behavior diffusion decreases and shows the characteristics of diminishing margins with the increase in the recovery rate μ of health food consumers. Figure 8 also demonstrates that a high heterogeneity of health food consumer networks indicates a low scale of the health food safety scare behavior diffusion.

5. Conclusion

The health food safety scare behavior diffusion is influenced by various factors. On the basis of the concept of complex network, this study extends the existing epidemic SIRS

(a)

(b)

(c)
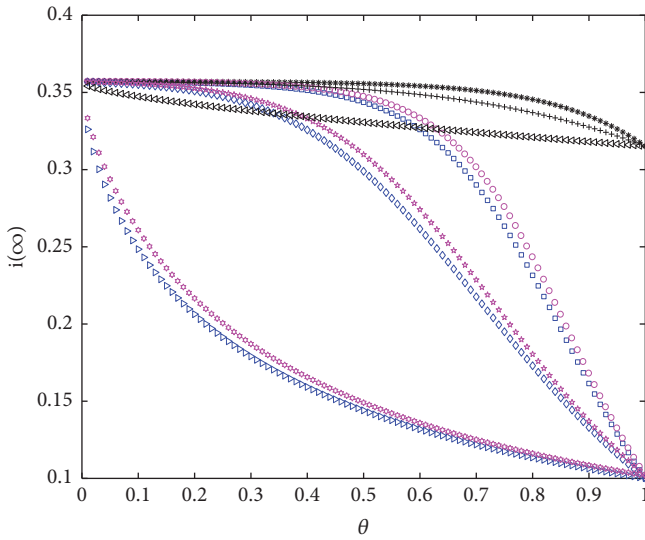
(d)

FIGURE 7: Continued.

(e)

(f)

Figure 7: Change rules of the scale $i(\infty)$ of the health food safety scare behavior diffusion. (a) Effects of the health food safety information transparency θ on $i(\infty)$ under different network structures and recovery rates μ (where $\lambda_0 = 0.001$, $\beta = 0.5$, $\eta_k = 0.7$, and $k = 1000$). (b) Effects of the degree k of health food consumers' ability to process information η_k on $i(\infty)$ under different network structures and recovery rates μ (where $\lambda_0 = 0.005$, $\beta = 0.5$, $\theta = 0.7$, and $k = 1000$). (c) Effects of the health food safety information transparency θ on $i(\infty)$ under different network structures and degree k of health food consumers' ability to process information η_k (where $\lambda_0 = 0.001$, $\beta = 0.5$, $\mu = 0.9$, and $k = 1000$). (d) Effects of the degree k of health food consumers' ability to process information η_k on $i(\infty)$ under different network structures and health food safety information transparency θ (where $\lambda_0 = 0.001$, $\beta = 0.5$, $\mu = 0.9$, and $k = 1000$). (e) Effects of the degree k of health food consumers on $i(\infty)$ under different network structures and degree k of health food consumers' ability to process information η_k (where $\lambda_0 = 0.001$, $\beta = 0.5$, $\mu = 0.9$, and $\theta = 0.1$). (f) Effects of the degree k of health food consumers on $i(\infty)$ under different network structures and health food safety information transparency θ (where $\lambda_0 = 0.001$, $\beta = 0.5$, $\mu = 0.9$, and $\eta_k = 0.01$).

model and establishes a heterogeneous model of the rate of health food safety scare behavior diffusion by considering health food safety information transparency and health food consumers' ability to process 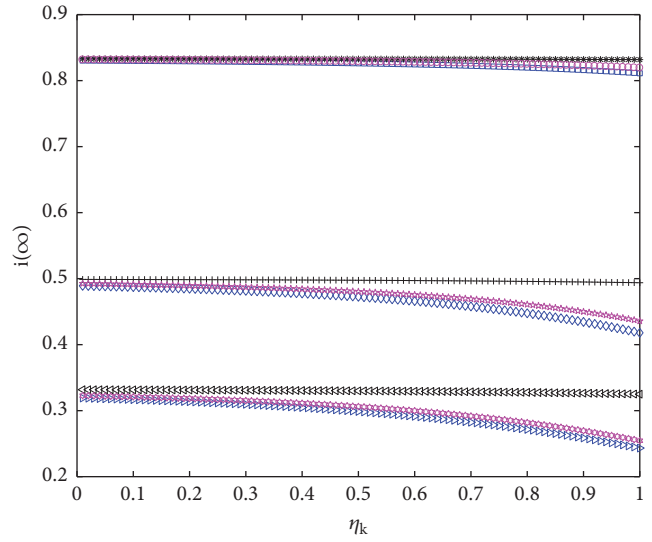information. The health food safety scare behavior diffusion and its influencing factors are analyzed theoretically using mean-field and network stochastic dominance theories. Mathematical simulation is performed to explore the effects of health food safety information transparency θ, health food consumers' ability to process information η_k, and heterogeneity of health food consumers' networks on the change rules of the threshold $\lambda_0 *$ of the rate and scale $i(\infty)$ of health food safety scare behavior diffusion. From the results, we have drawn the following main conclusions:

(1) Increases in the health food safety information transparency, health food consumers' ability to process information, and 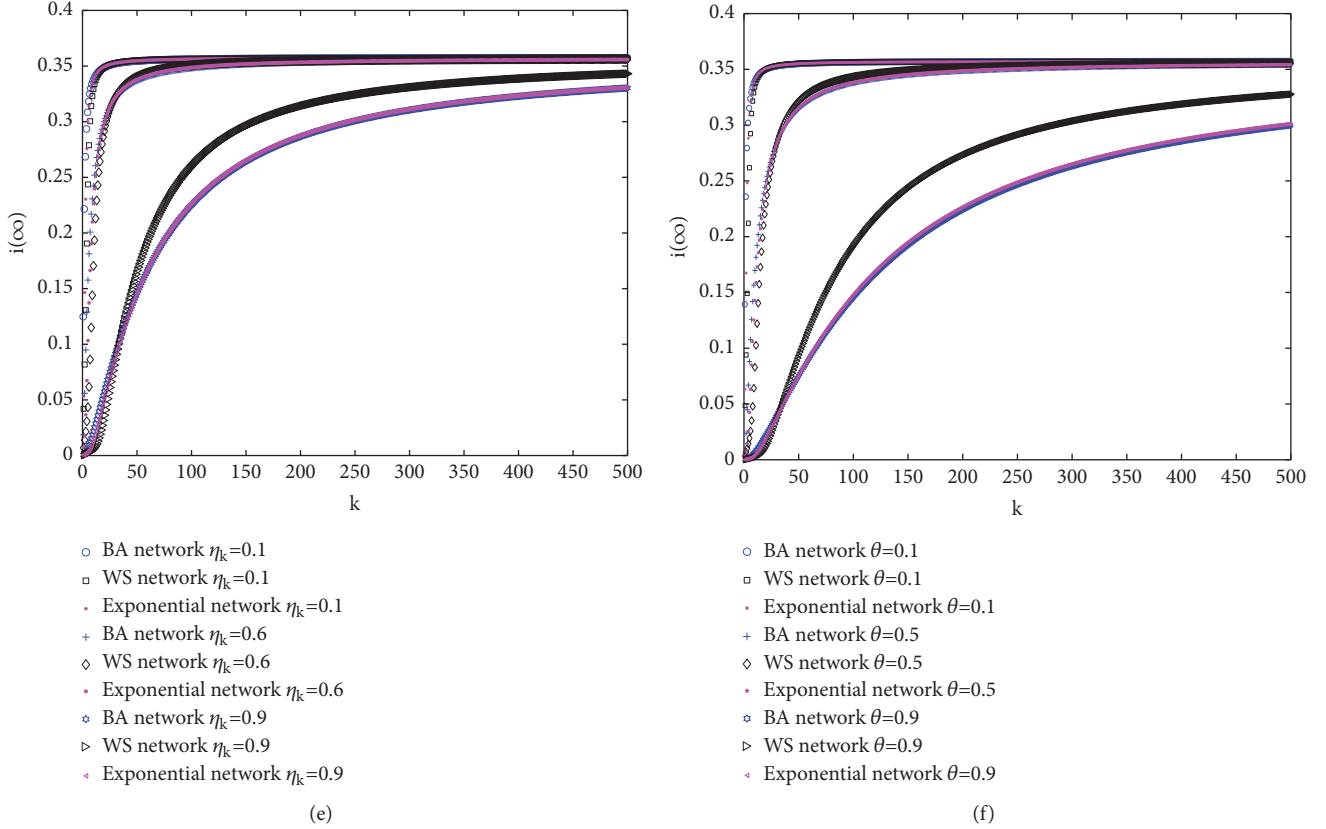recovery rate of health food consumers can increase the threshold of the rate of health food safety scare behavior diffusion. The health food safety information transparency

and recovery rate of health food consumers show marginal incremental rising characteristics in relation to the rate of health food safety scare behavior diffusion, whereas health food consumers' ability to process information reflects a marginal diminishing rising characteristic in relation to the rate of health food safety scare behavior diffusion.

(2) Increases in the health food safety information transparency, health food consumers' ability to process information, and recovery rate of health food consumers can decrease the scale of the health food safety scare behavior diffusion. The health food safety information transparency displays a marginal diminishing decreasing characteristic in relation to the scale of health food safety scare behavior diffusion, whereas the health food consumers' ability to process information and the recovery rate of the health food consumers indicate marginal incremental decreasing characteristics in relation to the scale of the health food safety scare behavior diffusion.
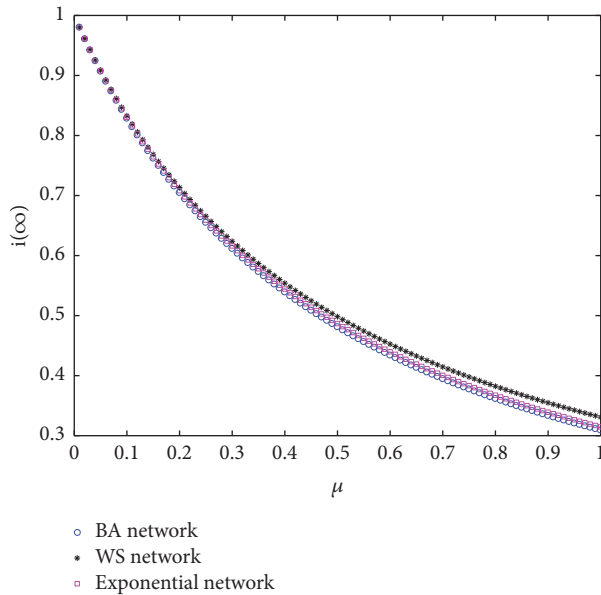
FIGURE 8: Effects of the recovery rate μ of health food consumers on the scale $i(\infty)$ of the health food safety scare behavior diffusion under three kinds of network structures (where $\lambda_0 = 0.005$, $\beta = 0.5$, $\theta = 0.75$, $\eta_k = 0.15$, and $k = 1000$).

(3) The network structure of health food consumers significantly affects the health food safety scare behavior diffusion. A high heterogeneity of the health food consumer network denotes a high threshold of the rate of health food safety scare behavior diffusion and a low diffusion scale.

These conclusions are theoretically and practically crucial to explaining the health food safety scare behavior diffusion. The information processing capacity of health food consumers can be improved by increasing the health food information transparency and guiding these consumers in increasing their health food knowledge. Effective recovery measures after health food safety accidents must be performed, and the structure of the health food consumer network must be changed to manage and control the health food safety scare behavior diffusion.

Data Availability

The method in this article is computer mathematical simulation. Numerical simulation analysis is the most effective way to test real-time dynamic data without a large number of empirical validations. The authors simulate to explore the characteristics and laws of the evolution of health food safety scare behavior diffusion by using Matlab2016b software. This paper does not have the data that can be obtained because they directly use the plot function of Matlab2016b software to make the images.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. Buchanan, "Understanding and Managing Food Safety Risks," *Journal of Food Safety*, vol. 16, no. 6, pp. 24–31, 2010.

[2] N. E. Piggott and T. L. Marsh, "Does food safety information impact U.S. meat demand?" *American Journal of Agricultural Economics*, vol. 86, no. 1, pp. 154–174, 2011.

[3] T. Chen, B. Ma, and J. Wang, "SIRS contagion model of food safety risk," *Journal of Food Safety*, p. e12410, 2017.

[4] T. Chen, L. Wang, J. Wang, and Q. Yang, "A Network Diffusion Model of Food Safety Scare Behavior considering Information Transparency," *Complexity*, vol. 2017, Article ID 5724925, 16 pages, 2017.

[5] T. Chen, L. Wang, and J. Wang, "Transparent assessment of the supervision information in China's food safety: A fuzzy-ANP comprehensive evaluation method," *Journal of Food Quality*, vol. 2017, Article ID 4340869, 14 pages, 2017.

[6] T. Li, J. C. Bernard, Z. A. Johnston, K. D. Messer, and H. M. Kaiser, "Consumer preferences before and after a food safety scare: An experimental analysis of the 2010 egg recall," *Food Policy*, vol. 66, pp. 25–34, 2017.

[7] D. Smith and P. Riethmuller, "Consumer concerns about food safety in Australia and Japan," *International Journal of Social Economics*, vol. 26, no. 6, pp. 724–742, 1999.

[8] J. M. E. Pennings, B. Wansink, and M. T. G. Meulenberg, "A note on modeling consumer reactions to a crisis: The case of the mad cow disease," *International Journal of Research in Marketing*, vol. 19, no. 1, pp. 91–100, 2002.

[9] M. L. Roehm and A. M. Tybout, "When will a brand scandal spill over, and how should competitors respond?" *Journal of Marketing Research*, vol. 43, no. 3, pp. 366–373, 2006.

[10] M. S. Park, H. N. Kim, and G. J. Bahk, "The analysis of food safety incidents in South Korea, 1998–2016," *Food Control*, vol. 81, pp. 196–199, 2017.

[11] N. Tanemura, N. Hamadate, and H. Urushihara, "The need for consumer science and regulatory science research on functional foods with health claims - What should we do to harmonize science and technology with society?" *Trends in Food Science & Technology*, vol. 67, pp. 280–283, 2017.

[12] J. A. Caswell and E. M. Mojduszka, "Using informational labeling to influence the market for quality in food products," *American Journal of Agricultural Economics*, vol. 78, no. 5, pp. 1248–1253, 1996.

[13] A. J. M. Beulens, D.-F. Broens, P. Folstar, and G. J. Hofstede, "Food safety and transparency in food chains and networks. Relationships and challenges," *Food Control*, vol. 16, no. 6, pp. 481–486, 2005.

[14] M. Mazzocchi, A. Lobb, W. B. Traill, and A. Cavicchi, "Food scares and trust: A European study," *Journal of Agricultural Economics*, vol. 59, no. 1, pp. 2–24, 2008.

[15] S. Cope, L. J. Frewer, J. Houghton, G. Rowe, A. R. H. Fischer, and J. de Jonge, "Consumer perceptions of best practice in food risk

communication and management: Implications for risk analysis policy," *Food Policy*, vol. 35, no. 4, pp. 349–357, 2010.

[16] A. P. J. Mol, "Governing China's food quality through transparency: a review," *Food Control*, vol. 43, no. 6, pp. 49–56, 2014.

[17] W. McKelvey, M. R. Wong, and B. Matis, "Letter Grading and Transparency Promote Restaurant Food Safety in New York City," *Journal of Environmental Health*, vol. 78, no. 2, pp. 46–48, 2015.

[18] A. Papadopoulos, J. M. Sargeant, S. E. Majowicz et al., "Enhancing public trust in the food safety regulatory system," *Health Policy*, vol. 107, no. 1, pp. 98–103, 2012.

[19] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[20] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[21] M. E. J. Newman, C. Moore, and D. J. Watts, "Mean-field solution of the small-world network model," *Physical Review Letters*, vol. 84, no. 14, pp. 3201–3204, 2000.

[22] K. Klemm, M. Á. Serrano, V. M. Eguíluz, and M. S. Miguel, "A measure of individual role in collective dynamics," *Scientific Reports*, vol. 2, p. 292, 2012.

[23] K. Musial, P. Bródka, and P. De Meo, "Analysis and applications of complex social networks," *Complexity*, vol. 2017, Article ID 3014163, 2 pages, 2017.

[24] P. Bródka, A. Chmiel, M. Magnani, and G. Ragozini, "Quantifying layer similarity in multiplex networks: a systematic study," *Royal Society Open Science*, vol. 5, no. 8, p. 171747, 2018.

[25] S. Gu, J. Johnson, F. E. Faisal, and T. Milenković, "From homogeneous to heterogeneous network alignment via colored graphlets," *Scientific Reports*, vol. 8, no. 1, 2018.

[26] A. Sharma, M. Kitsak, M. H. Cho et al., "Integration of Molecular Interactome and Targeted Interaction Analysis to Identify a COPD Disease Network Module," *Scientific Reports*, vol. 8, no. 1, 2018.

[27] S. E. Kessler, T. R. Bonnell, J. M. Setchell, and C. A. Chapman, "Social Structure Facilitated the Evolution of Care-giving as a Strategy for Disease Control in the Human Lineage," *Scientific Reports*, vol. 8, no. 1, 2018.

[28] J. de Jonge, L. Frewer, H. Van Trijp, R. Jan Renes, W. de Wit, and J. Timmers, "Monitoring consumer confidence in food safety: An exploratory study," *British Food Journal*, vol. 106, pp. 837–849, 2004.

[29] P. J. Williamson, "Sales and Service Strategy for the Single European Market," *Business Strategy Review*, vol. 3, no. 2, pp. 17–43, 2010.

[30] L. Austin, B. F. Liu, and Y. Jin, "How Audiences Seek Out Crisis Information: Exploring the Social-Mediated Crisis Communication Model," *Journal of Applied Communication Research*, vol. 40, no. 2, pp. 188–207, 2012.

[31] J. P. Quirk and R. Saposnik, "Admissibility and measurable utility functions," *Review of Economic Studies*, vol. 29, no. 2, pp. 140–146, 1962.

[32] J. Karamata, "Über einen Satz von Vijayaraghavan," *Mathematische Zeitschrift*, vol. 34, no. 1, pp. 737–740, 1932.

[33] J. Hadar and W. R. Russell, "Rules for ordering uncertain prospects," *American Economic Review*, vol. 59, pp. 25–34, 1969.

[34] G. Hanoch and H. Levy, "The efficiency analysis of choices involving risk," *Review of Economic Studies*, vol. 36, no. 3, pp. 335–346, 1969.

[35] M. Rothschild and J. E. Stiglitz, "Increasing risk. I. A definition," *Journal of Economic Theory*, vol. 2, pp. 225–243, 1970.

[36] P. C. Fishburn, "Convex stochastic dominance with continuous distribution functions," *Journal of Economic Theory*, vol. 7, no. 2, pp. 143–158, 1976.

[37] M. O. Jackson, *The Economics of Social Networks*, Cambridge University Press, Cambridge, UK, 2006.

[38] Y.-T. Bian, L. Xu, J.-S. Li, and X.-Q. Liu, "Dynamical evolution of trading behavior on anti-coordination game in complex networks," *China Finance Review International*, vol. 6, no. 4, pp. 367–379, 2016.

[39] V. Colizza and A. Vespignani, "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations," *Journal of Theoretical Biology*, vol. 251, no. 3, pp. 450–467, 2007.

[40] F. Zhang, L. I. Lu, and X. H. Yu, "Survey of transmission models of infectious diseases," *System Engineering Theory & Practice*, vol. 31, no. 9, pp. 1736–1744, 2011.

[41] S. Li and Z. Jin, "Dynamic modeling and analysis of sexually transmitted diseases on heterogeneous networks," *Physica A: Statistical Mechanics and its Applications*, vol. 427, pp. 192–201, 2015.

[42] R. Pastor-Satorras and A. Vespignani, "Epidemics and immunization in scale-free networks," in *Handbook of Graph & Networks*, S. ornholdt and H. G. Schuster, Eds., pp. 111–130, 2002.

[43] R. E. Lofstedt, "How can we make food risk communication better: Where are we and where are we going?" *Journal of Risk Research*, vol. 9, no. 8, pp. 869–890, 2006.

[44] J. He, X. Sui, and S. Li, "An endogenous model of the credit network," *Physica A: Statistical Mechanics and its Applications*, vol. 441, pp. 1–14, 2016.

Research Article

Examining the Intergovernmental and Interorganizational Network of Responding to Major Accidents for Improving the Emergency Management System in China

Pan Tang [1,2] **Haojia Chen,**[1] **and Shiqi Shao**[1]

[1]*School of Public Management/Emergency Management, Research Center of Emergency Management, Jinan University, Guangzhou, China*
[2]*Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong*

Correspondence should be addressed to Pan Tang; tangpan001@163.com

Since the SARS crisis in 2003, institutionalized emergency management systems have been established in each government level for improving inter-organizational collaboration in China. Major accidents require participation of public organizations affiliated with multiple government levels, and the lack of collaboration and coordination among the involved organizations within the critical time constraints during the response process is an existing problem. In this research, a case study of examining the intergovernmental and cross-sectoral collaboration for responding to a well-known oil pipeline explosion accident in China by a complex network method is conducted. The aim is to obtain managerial insights in improving the existing emergency management system in a centralized political-administrative context, such as China. A mixed method of data collection is applied to identify the participating organizations and to determine the interaction spanning organizational boundaries in both hierarchical and horizontal dimensions. An emergency response network is built and visualized for representing intergovernmental and interorganizational collaboration during the response process of the major accident by social network analysis (SNA) tools. The SNA indicators are used to measure quantitatively the network structure at the levels of the whole network, subnetwork, and node. The obstacles of achieving intergovernmental collaboration are found, and managerial suggestions for improving the existing emergency management system are provided. This research indicates that the Chinese government should pay attention to establishing and sustaining partnerships with private and nonprofit organizations and conduct a blend of hierarchical, market, and network principles in fostering collaboration for addressing major accidents. The public organizations in the local government level are shown to be more active than other participators in coordinating their response operations, and their capability should be emphasized for improvement. Additionally, the interactive relationships among specific emergency function groups and between the affected communities and organizations performing emergency command and coordination function should be strengthened.

1. Introduction

Major accidents always cause serious consequences to cities and overwhelm the capabilities of local governments. When emergency response extends above the local government to the provincial and central governments, public organizations affiliated with multiple government levels, as well as private and nonprofit sectors, are required to collaborate and interact with each other toward addressing the disastrous situation [1–3]. Effective response to major accidents depends

on intergovernmental and multiorganizational collaboration and networks [4, 5]. The dynamic and complex environment of rapidly evolving emergencies requires a different flexible approach than the traditional hierarchical mechanism. Legal authorities, responsibilities, resources, and information involved in responding to major accidents are currently dispersed among sectors affiliated with different governments in China. Such fragmentation makes accomplishing common objectives extremely difficult in emergency management. Lack of collaboration and coordination among the numerous

participants is an existing barrier that reduces emergency response effectiveness [6].

Interorganizational collaboration and networks for addressing emergencies has become widely accepted in practice [6, 7]. Since the SARS crisis in 2003, the Chinese government has been trying to establish an emergency management system [8]. For fulfilling the jurisdictional emergency management responsibility, each level of government designed and sustained an emergency management network consisting of governmental sectors, vertical management sectors, public companies, and institutions for improving collaboration among organizations with formal responsibilities and providing continuous emergency services to communities. Successful response to major accidents comprises rapid participation and collaboration among organizations from institutionalized emergency management networks in multiple government levels. As the accident breaks out, evolving and overwhelming capacities of local governments, governments at the higher levels activate the network once the disaster situations are evaluated to satisfy the specified conditions [9], and an intergovernmental network is formed for adapting to the dynamic situation. However, whether the established emergency management networks at multiple government levels can be induced to converge on the achievement of common objectives is an existing problem. Moreover, the function of each government level, as well as private and nonprofit sectors, should also be examined and explored.

Examining the intergovernmental network responding to major accidents provides an effective way to understand the collaborative process among the diversified participating organizations from system perspectives and helps to identify barriers of achieving successful emergency response collaboration. In addition, that will present implication to improve the existing emergency management system in China and to promote network-wide integration across government levels. During a response to major accidents, multiorganizational collaboration based on interorganizational relationships in both hierarchical and horizontal dimensions is essential toward addressing the disastrous situation [10]. In the hierarchical dimension, particularly in the context of China [11], interorganizational hierarchies specified in the political-administrative structure provide an important way to arrange response operations for facilitating collaboration among participants. Meanwhile, horizontal relationships among public organizations without formal hierarchical arrangement, as well as private and nonprofit sectors, are essential to improve interaction spanning organizational boundaries [6]. Furthermore, an appropriate emergency response network structure is essential for improving interaction among all the participating organizations, such as the development of common understanding of emergency situations and problems to be solved, the commitment to common incident objectives and plans, the extent to which all stakeholders are included in the process, and the recognition and management of interdependence among involved organizations [12–14].

The well-known oil pipeline explosion accident that occurred on November 22nd, 2011, in Qingdao City,

Shandong Province, provides a valuable opportunity to examine the intergovernmental and interorganizational collaboration in responding to special major accidents in the Chinese centralized political-administrative context. The investigation report of the State Council in China reveals that the failure of information sharing and lack of collaboration led to poor situation assessment and decision-making which directly caused disastrous consequences. Building on and contributing to the existing research on interorganizational collaboration and network analysis in emergency management, this research investigates the diversified participating organizations and varying levels of collaborative interaction across organizational boundaries in responding to the aforementioned major accident in China from network perspectives [15]. It aims to examine how public sectors affiliated with multiple government levels, as well as private and nonprofit sectors, interact and work together toward addressing major accidents in the context of the existing emergency management institutional arrangement in China. The following questions are examined and analyzed in China: (1) Which organizations participate in the emergency response process of addressing major accidents? (2) What are the structural characteristics of the emerged intergovernmental emergency response network? (3) What are the obstacles of achieving intergovernmental and cross-sectoral collaboration in this specific field from network structure perspectives? (4) How can the institutionalized emergency management system in China be improved?

The content analysis of multiple data sources was conducted to identify involved organizations and to determine diversified types of collaborative interaction across organizational boundaries in both hierarchical and horizontal dimensions in this empirical case. Social network analysis (SNA) concepts and tools are used to build, visualize, and analyze the complex emergency response network. The network structural properties are analyzed quantitatively using an SNA software tool at the level of the node, subset of nodes, and whole network. The key organizations involved in the response process are identified and discussed. In particular, block analysis is applied to present interactive relationships among groups of organizations with same attributes and shows the barriers of interorganizational collaborative response in the empirical case. Furthermore, the network analysis results are discussed and the possible improvement of the existing emergency management system in China is presented. The rest of this paper is organized as follows. Section 2 introduces the literature review of intergovernmental collaboration and relevant governance mechanisms and the emergency management network. The existing emergency management system of addressing major accidents in China is presented in Section 3 for providing the background of this research. The conceptual framework of intergovernmental and interorganizational networks is presented in Section 4. Section 5 presents the context of the empirical case, data sources, and research method. In Section 6, the emergency response network of the "11.22" oil pipeline explosion accident is built and visualized, and the network structures are measured and analyzed based on a SNA tool. The characteristics of the network of major accident response are

discussed and managerial suggestions in providing insight to improve the current emergency management system are provided in Section 7. Finally, Section 8 concludes this research.

2. Literature Review

The literature examining intergovernmental collaboration and the emergency management network provides the theoretical basis of this research and is discussed in this section.

2.1. Intergovernmental Collaboration and Governance. Intergovernmental collaboration across the governmental sector's boundaries and interorganizational collaboration among public, private, and nonprofit sectors are common in emergency management practice and are essential to achieve effective response to large-scale emergencies [16]. Each involved sector performs specific roles and responsibilities and should collaborate with others toward improving information communication, resource sharing, and action coordination. The mandated and emerging collaboration in addressing emergencies relies on the intergovernmental and cross-sectoral relationships [17]. On the one hand, the existing institutionalized interorganizational hierarchies specified in the political-administrative structure are employed to coordinate multiple organizations to achieve rapid collaboration when responding to major accidents [3]. On the other hand, responding to major accidents requires collaboration and coordination among public organizations without hierarchical relationships, as well as among public, private, and nonprofit sectors.

From the existing literature, addressing emergencies involving intergovernmental and cross-sectoral collaboration uses the combination of all the existing three governance mechanisms, including hierarchical, market, and network governance [3]. The first mechanism is the traditional model and relies on the chain of command, standardized procedures, and supervision. In the market mechanism, the supervision is less and the price is emphasized to sustain collaboration. Finally, the network mechanism reflects the horizontal reciprocal patterns of exchange and interrelationships among entities. The combination of multiple governance mechanisms necessitates the creation of a dynamic mixture governance mode to improve collaboration in emergency management.

2.2. Emergency Management Network. The widely existing intergovernmental and cross-sectoral collaboration forms the interorganizational network in the field of emergency management. The network provides an effective tool to define and understand the interdependence and varying levels of iteration among public organizations affiliated with multiple government levels, as well as private and nonprofit sectors [18]. The latest literature indicates that collaborative processes and structures work closely together in fostering effective interorganizational collaboration and determine the collaborative outcomes [19, 20]. In particular, the collaborative process among organizations and their relationships interacts with and is shaped by network structure arrangements. Therefore, appropriate emergency response network

structures are a necessary condition for facilitating successful interorganizational collaboration. In recent years, there has been rapid growing research on emergency management networks [21]. In particular, the empirical network research in emergency management using SNA as its methods is an emerging trend [11, 12, 22]. SNA refers to a range of methods of visualizing and analyzing interaction among nodes and provides tools to examine structural and relational patterns of social system and processes [11]. The existing research on emergency management networks concentrates on examining interorganizational collaborative processes at the node level, subset node level, and whole network level.

Despise rapid growth in network research in emergency management, theory construction, methodological rigor, and conceptual clarity are lacking. Second, the existing literature in this field is mainly on conducting research on interaction and collaboration among participating organizations in the western political-administrative context. The centralized political-administrative structure in China impacts the intergovernmental and multiorganizational collaboration deeply, which is different from the empirical emergency management network in most of the existing literature [11]. The existing research gap motivates us to examine the intergovernmental and multiorganizational collaboration in addressing special major accidents which involves all government levels in the institutionalized Chinese emergency management system.

3. Background of the Chinese Emergency Management System

This section introduces the existing emergency management system in China. It provides the institutional design for addressing major accidents, which is a typical interorganizational collaborative arrangement across multiple government levels.

After the SARS crisis in 2003, the Chinese government began to establish comprehensive emergency management systems for addressing both natural and man-made emergencies. For fulfilling the jurisdictional management responsibilities of addressing emergencies, each government established an emergency management network consisting of government sectors, departments of party committee, state-owned enterprises, and public institutions to improve collaboration among the participants in a jurisdictional area according to the Act on Addressing Emergencies of the People's Republic of China [23]. The specific emergency management responsibilities and roles of each participating organization are specified. First, an emergency management office (EMO) is set up at each government level to coordinate and manage the network activities of all the involved sectors as a network administrative organization [24]. An emergency management committee consisting of representatives from the main member organizations is established to address strategic-level problems, thereby leaving the daily operations to EMO. Second, each government sets up specific emergency command headquarters for coordinating involved organizations to address particular emergencies with higher
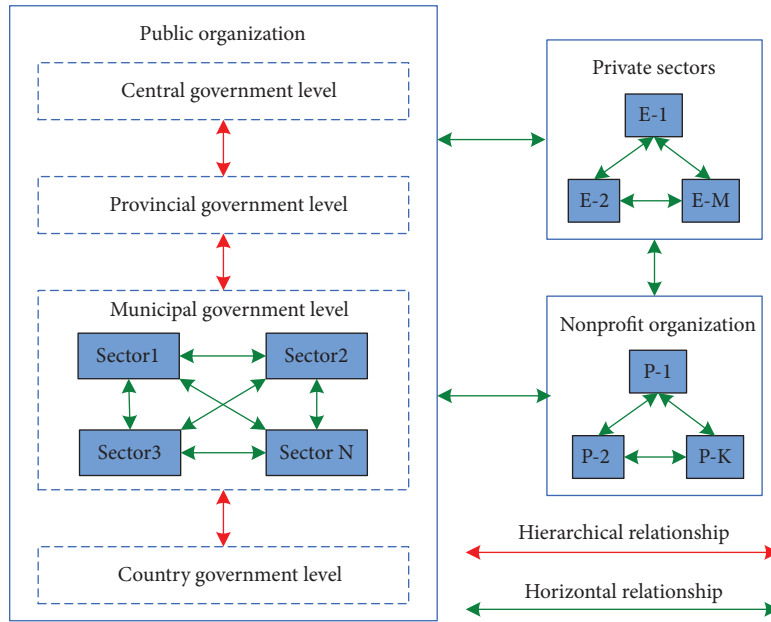
FIGURE 1: The conceptual framework of an intergovernmental and interorganizational network for responding to major accidents.

risk levels which may occur in the jurisdictional area. Furthermore, responsible organizations with similar capacities are arranged in an emergency function group to streamline the emergency services. A number of emergency operation plans have been developed for providing clear ground rules and processing transparency by arranging the response works of participators.

Hierarchically, the political-administrative structure in China consists of four governmental levels: the central, provincial, municipal, and county (or local) levels. The legal emergency management responsibilities of various government levels are specified according to the severity of consequence of emergencies in China. When a jurisdictional government evaluates that addressing the current emergency situation exceeds the capacity of subordinate governments and satisfies the conditions specified in the emergency operation plans, it should activate the established emergency management network in this government level, which is merged with the existing emergency response network for adapting to the disastrous situation. As the disastrous situation evolves and deteriorates, the emergency response network expands as governments in higher levels participate in addressing the emergencies. The layered emergency management institutional arrangement allows for appropriate response scale for each specific emergency.

In the current time, the emergency management network is designed and sustained independently within different government levels. How to leverage and integrate emergency management networks across multiple government levels within critical time constraints toward addressing major accidents together is an existing problem. In addition, the actual emergency response network should take into account the emergent characteristics and include a broad enough spectrum of participators for adapting to the complex emergency situation [25, 26].

4. The Conceptual Framework of the Intergovernmental Emergency Response Network

During major accidents, an emergency response network consisting of diversified participators emerges and sustains adapting to the disastrous situation. While performing different types of operations, the broad range of participating organizations interact and collaborate with one another across the organizational boundaries, which is essential for achieving successful emergency response. Actually, the involved interorganizational interactions reflect an intriguing mixture of collaboration in both hierarchical and horizontal dimensions. In this section, a conceptual framework of an intergovernmental and interorganizational network is presented for examining the intergovernmental and cross-sectoral collaboration of addressing major accidents as shown in Figure 1. The participating organizations, both the hierarchical and horizontal interorganizational relationships, as well as the involved governance mechanisms [27] of promoting interorganizational collaboration, are discussed. The details are listed as follows.

4.1. Hierarchical Relationships and the Governance Mechanisms. In the context of China, the hierarchical relationships are mainly the interorganizational hierarchies specified by the institutionalized political-administrative structure [28] during emergencies. First, the affiliation relationships between the jurisdictional government and its subordinate sector, such as governmental sectors, state-owned companies, and public institutions, are typical hierarchical relationships. Second, the administrative supervision relationships between governments at various levels and business guidance relationships among the governmental sectors in different government levels belong to this type of

interorganizational relationships. The top-down administrative system consists of the specified interorganizational hierarchical relationships, and formal authority is extended along with the hierarchies.

On the basis of interorganizational hierarchies, organizations with authorities direct and control multiple subordinate sectors to perform specific operations and resolve possible conflicts in a timely manner. Hierarchy mechanism is an essential governance mechanism to coordinate multiple organizations with these types of interorganizational relationships. The jurisdictional government mobilizes all the subordinate sectors to participate in the response process and achieves rapid coordination at a government level. Meanwhile, among various government levels, public organizations at a lower level are the first to respond to emergencies and report the emergency situation to those at the higher level. Then, the government at the higher level provides external assistance to local governments, as well as coordinates and supervises all subordinate sectors toward addressing a disastrous situation. In China, the provincial and municipal governments play essential roles as a conduit for hierarchical collaboration between central and local governments in large-scale emergency response (Lu, 2016).

4.2. Horizontal Relationships and the Governance Mechanisms. Horizontal relationships are those between public organizations without hierarchical relationships, interjurisdictional relationships, and cross-sectoral relationships between public sectors and those from market and society. The interjurisdictional relationships are those among different jurisdictional governments, whereas the other horizontal relationships focus on the broader relationships among organizations without hierarchical relationships, such as those among sectors affiliated with the same jurisdictional government and those among public, private, and nonprofit sectors. Horizontal relationships reflect interaction and exchanges among autonomous organizations, which stress autonomy, partnership, and voluntary agreements. Responding to major accidents creates an atmosphere to develop and improve horizontal relationships, wherein numerous organizations share the common risks and responsibilities. From the institutional arrangement of each government in China, the EMO establishes and sustains the horizontal relationships among organizations with formal emergency management responsibilities for improving information sharing and broad collaboration among all involved sectors in the jurisdiction as discussed in Section 3. The emergency operation plans of addressing accidents specify the formal responsibilities of each involved organization and group them into several emergency function groups for providing streamlined emergency service to impacted communities [29]. Participating organizations affiliated with the same emergency function group establish and maintain horizontal relationships by programing emergency service procedures and excising together toward improving joint decision-making and implementation. Finally, jurisdictional governments at various levels frequently collaborate with one another by signing mutual aid agreements to build interjurisdictional partnerships for sharing information and resource or coping with shared risks efficiently which may spread across administrative boundaries.

Similar to hierarchical collaboration, interorganizational, interjurisdictional, and cross-sectoral collaboration and interaction in the horizontal dimension are also essential to improve effective response in actual emergency situations. However, collaboration across organizational boundaries based on horizontal relationships mainly relies on a network mechanism and market mechanism to arrange the operations of involved sectors. The later one allows participants to use their individual resource to achieve their self-interest, where the former one emphasizes shared value, trust, and consensus in improving interorganizational collaboration and coordination [30]. In particular, network and market mechanisms are essential to mobilize resources beyond the government system and integrate capacities of various sectors from the market and the society toward addressing major accidents.

As per the above discussion, responding to major accidents involves intergovernmental and cross-sectoral interactions and requires effective integration of hierarchical and horizontal relationships that complement each other [3]. Hierarchical, market, and network mechanisms are mixed and embedded in their forms of interorganizational relationships to regulate the behaviors of involved organizations toward improving collaboration and coordination among multiple levels of governments, as well as private and nonprofit sectors. How to leverage and combine each activated emergency management network in various government levels from both hierarchical and horizontal dimensions to regulate diversified interorganizational interactions toward addressing major accident collaboration is an existing problem, particularly in the centralized political-administrative structure in China.

5. Context, Data Source, and Research Method

In this research, a case study [31] is conducted to examine the involved intergovernmental and cross-sectoral collaboration during the process of responding to the well-known "11.22" Donghuang oil pipeline explosion accident in China from network perspectives. In this case study, the analysis unit is the participating organizations, and various types of interactions across organizational boundaries are recorded for building the actual emergency response network which emerged rapidly after the occurrence of the explosion accident. The case is summarized to provide the context of this study. Furthermore, the data sources and research method are also introduced.

5.1. Context of the Study. At 10:25 pm on November 22nd, 2011, an oil spill occurred at the Donghuang petroleum transmission pipeline, which is located within a highly urbanized and coastal area in the Economy and Technology Development Zone of Qingdao City, Shandong Province, in eastern China. The leaked oil flowed into the municipal pipe network and caused a huge explosion. This accident killed 62 persons and injured 136 others. Furthermore, it caused serious damage to the surrounding construction structures,

and the water, electricity, heat, and gas supplying systems were destroyed to varying degrees. The leaked oil flowed into the nearby sea through the municipal pipes, caused serious environmental pollution in the coastal region, and were burnt by the explosion, which threatened the safety of the multiple oil tanks in this city. The direct economic losses amounted to nearly 750 million RMB.

The consequences of this accident are evaluated to satisfy the conditions of special major accidents. According to the institutional arrangements of the emergency management system in China as introduced in Section 3, as the accidents occurred and evolved, all of the local, municipal, provincial, and central governments activated emergency management networks in each level in sequence and provided support to the on-scene emergency response operations. Meanwhile, private and nonprofit organizations also participated in the response and recovery process of this accident. Furthermore, a number of emergency response functions are involved, such as firefighting, search and rescue, medical care, cleaning of the oil spill pollution in the coastal and land regions, resettlement of the victims, and repairing of water-, electricity-, gas-, and heat-supplying systems. As a result, responding to this accident involved a broad array of organizations with specific responsibilities and capabilities.

The report [32] issued by the investigation team of the State Council reveals the failure of intergovernmental and interorganizational interaction and collaboration during response to this explosion accident. In particular, poor information reporting and sharing among sectors affiliated with local government and between private and public sectors is the main problems to be identified. Furthermore, the situation assessment and response behaviors of participating organizations are impacted by the interorganizational interactions and caused the disastrous consequences. Therefore, this real-world accident provides an opportunity to examine the intergovernmental and cross-sectoral collaboration by building and analyzing the emerged emergency response network in the context of the Chinese centralized and hierarchical political context. In this research, analyzing the emergency response network of the "11.22" oil pipeline explosion by SNA tools provides an effective way to identify the barriers of integrating all the formal established emergency management networks in each government level and to examine the emerging characteristics of the response process.

5.2. Data Sources. In this research, mixed methods of data collection are conducted to identify the involved participating organizations and interorganizational interactions at varying levels during the response process to the "11.22" oil pipeline explosion accident. All the involved participating organizations are identified and coded by network nodes, while the diversified interorganizational interactions are recorded as network ties for building the emergency response network. First, content analysis of multiple data sources, such as the activated emergency operation plans during response process, newspapers, situation reports, and other reports from official websites and Weibo, was conducted. The detailed explanation of each data source is listed as below.

(1) The Activated Emergency Operation Plans. Emergency operation plans describe the legal roles and responsibilities of each organization and provide guidance for emergency management. The jurisdictional governments at the local, municipal, provincial, and central levels activated relevant emergency operation plans during the "11.22" oil pipeline explosion accident for addressing disastrous situations. Content analysis of these documents was conducted to identify the candidates of participating organizations. As a result, a preliminary list of participating organizations was established.

(2) Official Accident Investigation Report. After the explosion accident occurred, the State Council of China set up an official investigation team and issued an investigation report on January 11th, 2014. This official document described the detailed emergency response process and provide a data source with high reliability, which specifies the main participating organizations and interactions spanning organizational boundaries [32].

(3) Situation Reports from Newspapers, Official Websites, and Weibo. This research also collected related reports from Qingdao Daily, Qingdao Evening Daily, and Qingdao Morning Daily. Related electronic media reports from the official website and Weibo of the provincial, municipal, and local government sectors were selected as an authoritative source of trusted information to track the emergency response process and to collect data of network nodes and ties.

Second, in-depth interviews of key informants were conducted to complement with the primary data sources. Numerous public organizations at the municipal and county government levels participated in the emergency response process in this case. Two emergency managers from the Emergency Management Office of Qingdao City, which are responsible for managing the emergency management network at the municipal government level, were interviewed. In addition, six managers from sectors affiliated with the Government of Qingdao Economic and Technological Development Zone were interviewed to investigate the participating organizations and the involved types of interorganizational interactions. The first-hand data complements with the aforementioned second-hand data sources.

5.3. Research Method. For examining the intergovernmental and cross-sectoral collaboration among all the participating organizations, network analysis was conducted to identify the structural patterns of their relationships and to analyze the effects of the network structure on the response behaviors of participators [33]. This research applied the SNA software tool Netminer [34] to model and visualize the involved organizations and their interorganizational relationships in both hierarchical and horizontal dimensions during the responding process to the "11.22" explosion accident. Meanwhile, the SNA indicators were applied to measure the network structure at the levels of the node, subset of nodes, and whole

network. Furthermore, how the network structure characteristics affect behaviors of involved organizations was discussed, and the managerial implications were provided. The involved network metrics, such as network density, centrality analysis, and structural block analysis, are focused on and introduced as follows.

Network density indicates the degree to which a network is cohesive. The SNA measures, such as density and centralization, can be applied to depict the network structural properties. The density is defined as the number of existing ties between organizations, with respect to the maximum number of possible ties.

Centrality measure is used to identify the key actors in the network or describe the nodes' position or roles. This indicator reveals interesting characteristics about the network. The centrality analysis includes degree centrality, betweenness centrality, closeness centrality, and effect centrality [33]. Degree centrality analysis can be performed to identify which organization interacts directly with another in the network. Betweenness centrality is an indicator of the extent to which a network actor locates within the shortest distance between a pair of nodes in the network [33]. Its value ranges from 0 to 1. The higher betweenness centrality value for a network actor indicates that it can better control the information communication of other nodes in the network. The effect centrality of a given node is the measure of the effect strength from this node to all the other ones through every path between them. It fully captures the concept of embeddedness because it considers both direct and indirect links. Therefore, we believe that this concept is a more accurate indicator of the extent to which an organization is embedded in the network structure comparing with the degree centrality, which only considers the local area of the network.

In social network analysis, all the nodes can be divided into different exclusive subsets, called positions. Structural block analysis can be employed to describe the interactive relationships among positions [33]. The produced block model is an abstract representation of the entire network. For each pair of positions, the structural block analysis reports whether the ties are present or absent within or between those pairs. Moreover, the density of a given interactive relationship is the ratio of the number of present lines to the maximum number of possible ones.

6. Building and Analyzing an Emergency Response Network for Major Accidents

Responding to major accidents is a highly complex process that involves numerous response tasks that should be completed to achieve common incident objectives. An emergency management network consists of diversified organizations performing different tasks and interacting with one another. The interactions and exchanges spanning organizational boundaries are of varying levels and difficult to identify, thereby presenting obstacles to build the emergency response network. The content analysis of the aforementioned data sources was conducted to identify interorganizational activities for building an emergency response network of addressing the "11.22" oil pipeline explosion accident. In fact,

it is an intergovernmental, cross-sectional, and interjurisdictional network for adapting to the complex disastrous situations. Network analysis based on the SNA tool Netminer [34] is employed to visualize, measure, and decipher the network. The involved interorganizational collaborative process is analyzed and understood from network structural perspectives. This section introduces the approach of identifying response organizations and defining the boundary of the emergency response network. Also, how to evaluate interorganizational collaborative relationships in the institutional context of China by tracking the interaction among organizations during emergencies is presented. Finally, the emergency response network is visualized and analyzed for examining characteristics of the involved intergovernmental and cross-sectoral collaborative process.

6.1. Identifying the Response Organizations. Identifying organizations that participated in the actual emergency response process of the "11.22" oil pipeline explosion accident is the first step to build the emergency response network. Major accidents cause disastrous consequences and require a broad range of response organizations to perform diversified tasks toward achieving common objectives. The emergency response process increasingly involves close interactions across an array of sectors and different government levels [35]. Moreover, the actual emergency response tends to involve an unpredictable set of organizations rather than follow expectations in the documented emergency operation plans [36]. Therefore, identifying all participating organizations and specifying network boundary remain as dilemmas in the field of emergency response [37].

The operational criteria for inclusion in this research requires member organizations to perform tasks or to provide resources for achieving common objectives of addressing the studied accident. First, the candidate participating organizations were identified by content analysis of the activated emergency operation plans at each government level, and then the preliminary response organization list was formed. These organizations have formal emergency management responsibilities, and most of them are from the public sector. However, the emergency response demonstrates the emergent properties [25]. Most of the private and nonprofit organizations were not listed in the emergency operation plans but actually participated in the emergency response and contributed to the achievement of incident objectives. Thus, all other data sources were utilized to verify whether organizations in the preliminary list actually responded to disastrous situations and to identify other organizations which satisfy the abovementioned operational criteria of network members. Several new organizations were added to the initial organization list. Finally, a total of 209 organizations which participated in the response process to the "11.22" oil pipeline explosion were identified as shown in Table 1. Each organization was recorded as a separate entity, and multiple attributes were encoded, such as organization number, name, organizational type, responsibilities, and the main performed tasks. All the response organizations were classified as belonging to either the public, private, or nonprofit sector. In addition, the public sectors affiliated with

Table 1: The organizations responding to the "11.22" oil pipeline explosion.

Organization type	Number	Percentage
Public organization		
Central level	8	3.8%
Provincial level	10	4.8%
Municipal level	34	16.3%
County level	37	17.7%
Private organization	65	31.1%
Nonprofit organization	55	26.3%
All organizations	209	100%

various government levels consist of government sectors, state-owned companies, and public institutions.

6.2. Evaluating Interorganizational Relationships. During emergency response, all the participating organizations interact with one another by conducting diversified activities spanning an organization's boundaries. Interorganizational relationships describe which participating organizations interact with one another and provide abstract representation of multiple types interorganizational interaction, such as issuing orders and making commitments; information reporting; providing resource, service, and expertise; information sharing; joint decision-making by attending meetings; sharing resource and knowledge; and working together when performing common tasks. In this research, the interorganizational relationships are determined by identifying interactive activities across the organization's boundaries. The content analysis of the aforementioned data sources in Section 5.2 which records the emergency response process was conducted to evaluate the interorganizational relationships.

As discussed in Section 4, interorganizational relationships among all the participators consist of hierarchical and horizontal relationships. Hierarchical relationships are established and sustained before emergencies and reflect the hierarchy mechanism. As formal network ties, the preliminary list of hierarchical relationships among participating organizations was identified according to the institutional hierarchical arrangements in the emergency management system in China. Furthermore, each hierarchical relationship in the list is verified by identifying interactive activities of issuing orders and making commitments and reporting information among the participating organizations in the actual response process. Only those representing actual interactions spanning organizational boundaries were retained in the list. On the other hand, interorganizational relationships in the horizontal dimension are conceptualized as a representation of interactive activities among the identified participating organizations without hierarchical arrangements. Most of the emergent interorganizational interactions belong to this type. A context analysis of data sources was conducted to identify collaborative activities across organizational boundaries, such as issuing orders, making commitments, information reporting and reception, for determining whether there exist horizontal relationships among the participating

organizations. Unlike hierarchical relationships, horizontal ones are determined by formal collaborative arrangements and emergent interorganizational interaction and exchanges during emergencies.

In this research, a total of 141 hierarchical relationships and 521 horizontal relationships among the 209 participating organizations were identified and verified. The hierarchical relationships function as stronger network ties compared with the horizontal ones. Both types of relationships complement each other and simultaneously improve collaboration and coordination among all the participating organizations by applying different governance mechanisms as discussed in Section 4.

6.3. Visualizing the Emergency Response Network. Effective response to major accidents depends on the integration of all the involved partners, which interact with one another during emergencies. All response organizations should understand their respective roles and responsibilities, as well as how to complement each other toward achieving common incident objectives. The emergency response network of the "11.22" oil pipeline explosion accident is built and shown in Figure 2 based on the identified participating organizations and interorganizational relationships. This interorganizational network consists of 209 organizations, with 141 links representing hierarchical relationships and 521 links representing horizontal ones. This actual emergency response network comprises a combination of organizations from local, municipal, provincial, and central government levels, as well as private and nonprofit organizations. In addition, interorganizational relationships in both hierarchical and horizontal dimensions interweave with each other, and the underlying governance mechanisms complement each other contributing to improve the coordination and collaboration among the involved response organizations [38]. The whole network is integration of all the activated emergency management networks in each government level, as well as the emergent nodes and dyadic ties.

6.4. Analyzing and Deciphering the Network Structure. The emergency response network of the "11.22" oil pipeline explosion accident exhibits interorganizational interactions in both hierarchical and horizontal dimensions; it shapes the actual collaborative process among all the response organizations in the Chinese institutional environment of emergency management. In this section, the quantitative metrics in SNA are used to analyze the emergency response network. The characteristics of the emergency response network are examined at the levels of the whole network, subnetwork, and node. In addition, how the network characteristics impact the behaviors of these participating organizations is analyzed and discussed.

6.4.1. Characteristics of the Whole Network. The density of the aforementioned emergency response network is 0.031, and its average degree is 6.396. Those network analysis results indicate that one organization interacts with an average of six organizations. Network density also shows the sparse characteristic of the emergency response network.
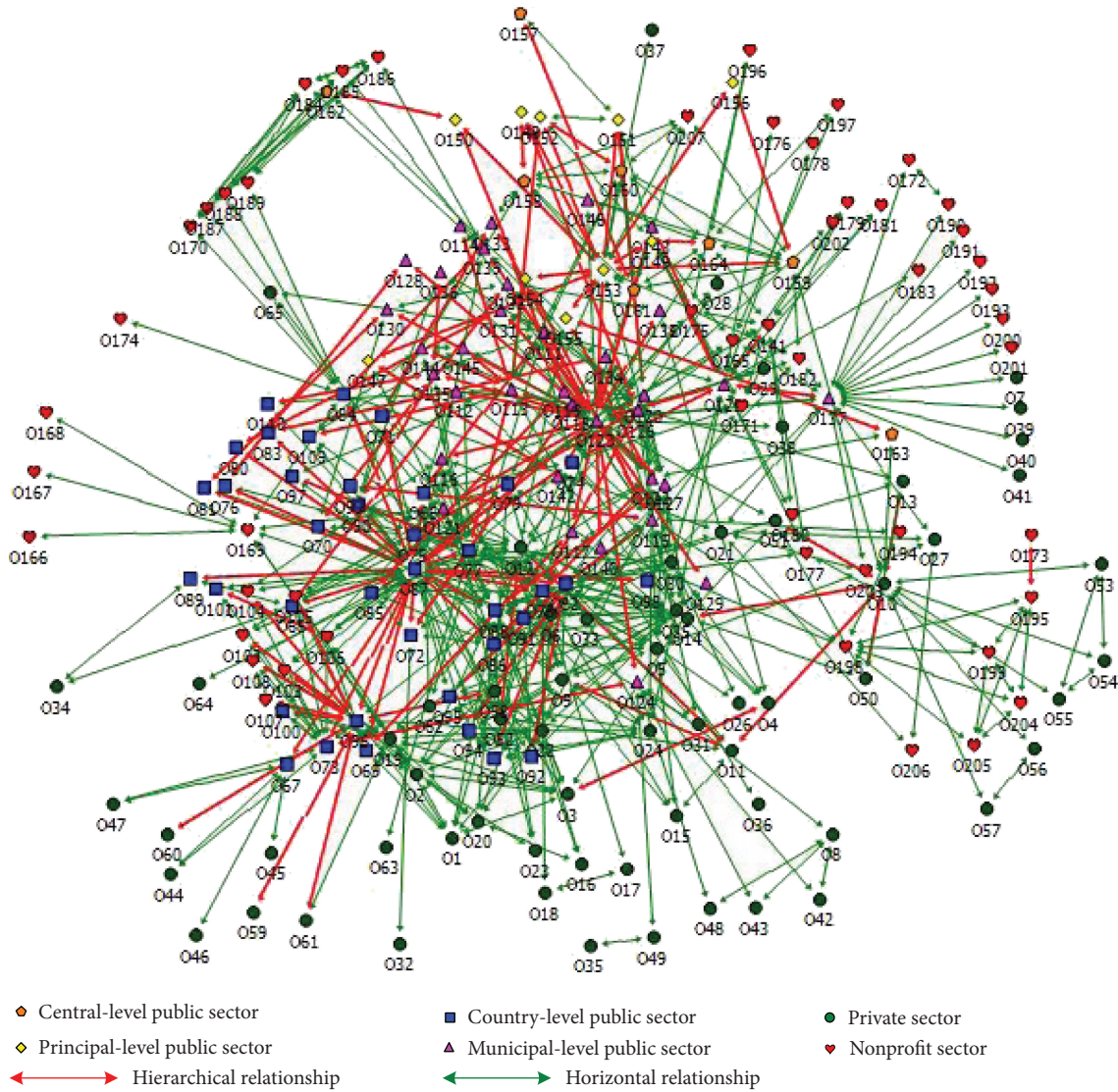
FIGURE 2: Emergency response network of the "11.22" oil pipeline explosion.

6.4.2. Centrality Analysis. Centrality analysis is used to analyze the embeddedness of nodes in the network. As an important network metric, it determines access to and control over resources and information in the network structure. Analyzing how a response organization is embedded into the whole emergency response network is an effective way to understand its behaviors during emergencies.

(1) Degree Centrality Analysis. Figure 3 presents a concentric map of degree centrality. The greater the degree centrality value of the response organization is, the closer its location to the center of the map. Table 2 shows the 10 response organizations with the highest degree centrality value. Figure 3 indicates that most of the private sector and non-profit organizations is located at the periphery. The response organizations in the core are mainly the public organizations. In particular, public organizations are mainly from the municipal and district government levels. The municipal

and district governments and sectors affiliated with them are the main response organizations in the actual emergency response process of this major accident.

Table 2 shows that EMOs of Huangdao District (O87) and Qingdao District (O123) keep interaction with most organizations during emergencies. In the top 10 organizations with high degree centrality value, 6 are from the district government level, whereas the remaining 4 are municipal government level ones.

(2) Betweenness Centrality Analysis. Table 3 shows the top 10 organizations with the highest betweenness centrality. With the exception of response organization O10, all the organizations are public organizations at the district, municipal, and provincial government levels. In particular, the EMOs of the Qingdao Municipal Government (O123) and Huangdao District Government have higher betweenness centrality value than the other organizations. This result indicates that
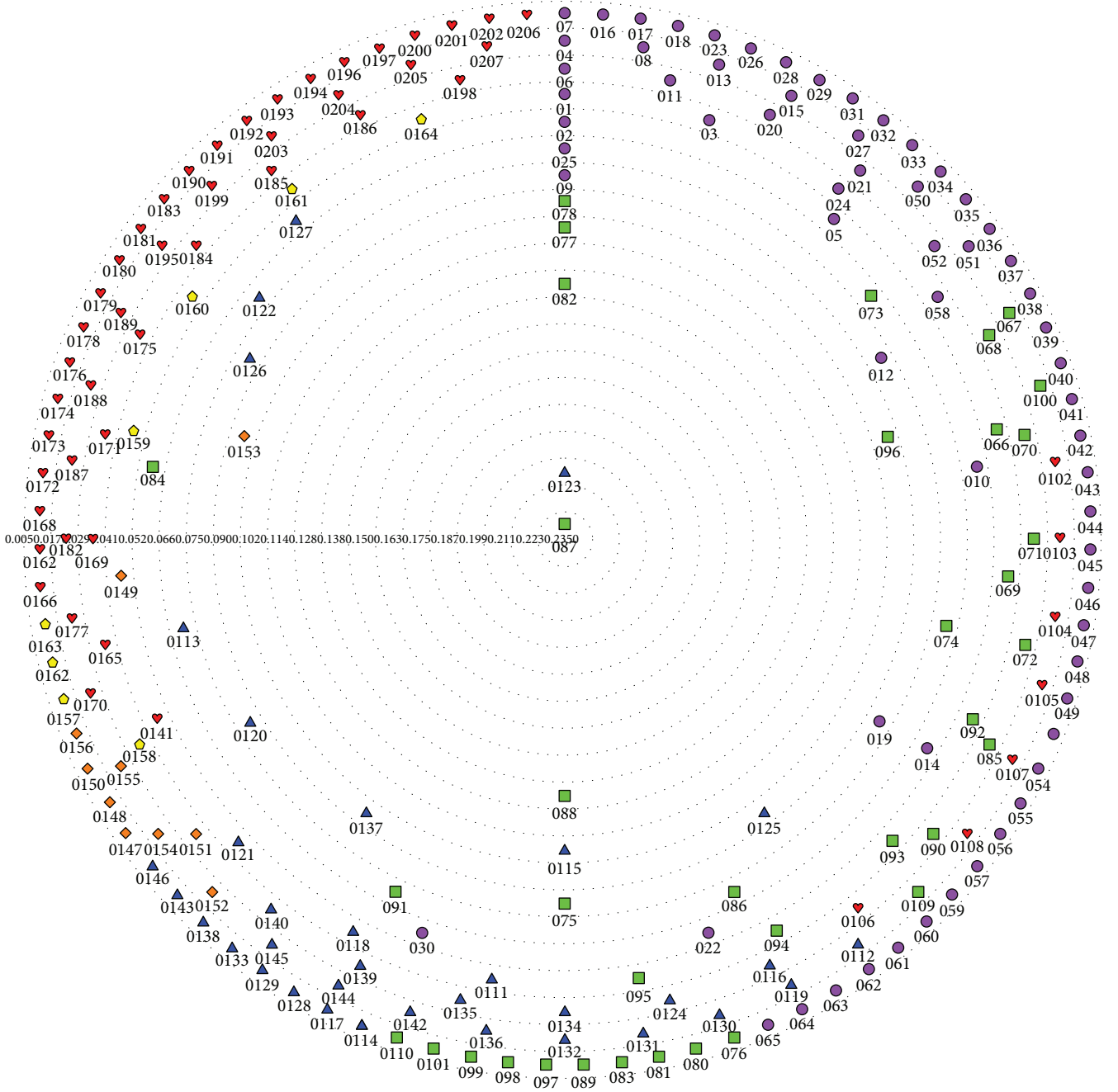
FIGURE 3: Concentric map of the response organizations.

the EMO at the municipal and district levels of the government are the information hubs, which play essential roles in collecting and disseminating information.

(3) Effect Centrality analysis. The weight parameter is set to be 0.6 in the effect centrality analysis in this study, which defines the transmitted effect of indirect links on the direct link. The ten response organizations with the highest effect centrality are listed in Table 4. As shown, five response organizations are public organizations at the district government level, four organizations are from the municipal government level, and only one is from the provincial government level. The effect centrality analysis shows that

public organizations have a stronger impact on the whole network.

6.4.3. Structural Block Analysis. In this section, the whole characteristics of the emergency response network are examined. Each position represents a mutually exclusive subset of organizations with the same organizational attribute. The block model presents interactive ties among all the positions, which provides an effective way to examine characteristics of the emergency response network.

(1) Block Analysis of Organizational Types. As discussed in Section 2, each jurisdictional government establishes and

TABLE 2: Ten organizations with the highest degree centrality.

Rank	Organization code	Organization name	Degree centrality
1	O87	EMO of the Huangdao District Government	0.2476
2	O123	EMO of the Qingdao Municipal Government	0.2184
3	O88	Huangdao District Public Works Administrative Departments	0.1359
4	O82	Huangdao District Firefighting Bureau	0.1262
5	O77	Huangdao District Police Bureau	0.1068
6	O115	Qingdao Municipal Firefighting Bureau	0.1068
7	O78	Huangdao District Environment Protection Bureau	0.0971
8	O96	Huangdao Street Office	0.0971
9	O125	Qingdao Municipal Health Bureau	0.0971
10	O137	Qingdao Municipal Central Blood Station	0.0971

TABLE 3: Ten organizations with the highest betweenness centrality.

Rank	Organization code	Organization name	Betweenness centrality
1	O123	EMO of the Qingdao Municipal Government	0.3132
2	O87	EMO of the Huangdao District Government	0.2707
3	O137	Qingdao Municipal Central Blood Station	0.1213
4	O88	Huangdao District Public Works Administrative Departments	0.0961
5	O125	Qingdao Municipal Health Bureau	0.0642
6	O115	Qingdao Municipal Firefighting Bureau	0.0614
7	O124	Qingdao Municipal Business Bureau	0.0597
8	O84	Huangdao Municipal Education Management Bureau	0.0564
9	O153	EMO of the Shandong Principal Government	0.0558
10	O10	China Petroleum and Chemical Corporation	0.0515

TABLE 4: Ten organizations with the highest effect centrality.

Rank	Organization code	Organization name	Effect centrality
1	O123	EMO of the Qingdao Municipal Government	0.0233
2	O87	EMO of the Huangdao District Government	0.0230
3	O137	Qingdao Municipal Central Blood Station	0.0210
4	O125	Qingdao Municipal Health Bureau	0.0124
5	O88	Huangdao District Public Works Administrative departments	0.0102
6	O96	Huangdao Street Office	0.0099
7	O153	EMO of the Shandong Principal Government	0.0098
8	O77	Huangdao District Police Bureau	0.0096
9	O115	Qingdao Municipal Firefighting Bureau	0.0092
10	O82	Huangdao District Firefighting Bureau	0.0089

sustains an emergency management network for facilitating collaboration among involved organizations in jurisdictional areas. In this research, the response organizations are grouped into six groups, such as public organization groups at the central, provincial, municipal, and county government levels, a private organization group, and a nonprofit organization group. Addressing "11.22" oil pipeline explosion requires effective collaboration and coordination among all of these groups. However, the motivations, institutional logic, and accountability mechanisms of response

organizations affiliated with these groups are different, which impacts the behaviors and interorganizational interaction during emergency response [28].

Figure 4 presents the interactive relationships among and within blocks representing the above groups. The lines represent interactive relationships and the width and color of the lines represent the density of network ties, so that a wider and darker line indicates a higher density of ties among or within groups. As shown in Figure 4, public organizations affiliated with different government levels prefer to interact
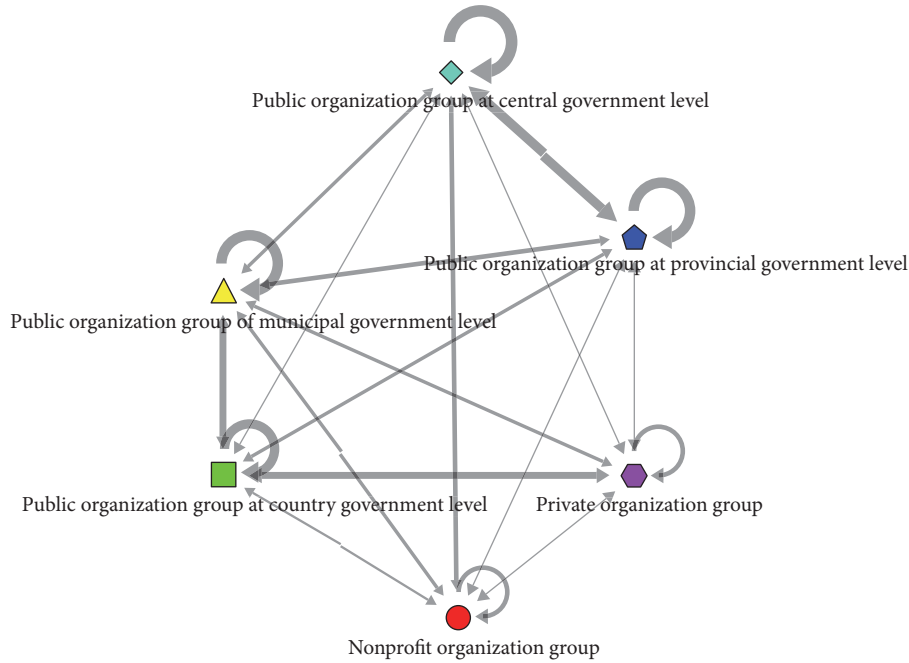
Figure 4: Block model of response organization type.

with other ones in the same group. The phenomena indicate that the established emergency management networks in each government level improves interorganizational collaboration among public organizations effectively. In addition, the analysis result indicates that public organizations at the same government level pay more attention to collaborating with each other in the same group than with private and nonprofit organizations and reflects that the cross-section horizontal relationships among public, private, and nonprofit organizations are not established, maintained, and nurtured at the current time. The existing intergovernmental interactions are mainly between jurisdictional governments at adjacent levels. In fact, that reflects the hierarchical structure among public organizations affiliated with different government levels and shows that the political-administrative system of the Chinese government plays essential roles in facilitating intergovernmental collaboration and coordination during this major accident. The private organizations tend to collaborate with organizations within the same group and public organizations at the county government level, which is at the bottom of the Chinese government system. The interactive relationships between private organizations and public organizations at the central, provincial, and municipal government levels are lacking. Finally, close interactions and links between nonprofit sectors and all other response organizations are lacking. That reduces the collaboration level and emergency response effectiveness.

(2) Block Analysis of Emergency Function Groups. Similarly, the response organizations can also be grouped into multiple exclusive subsets representing different emergency functions [16]. Response organizations affiliated with an emergency function group have specific capabilities for providing a specific emergency service. Moreover, emergency services provided by various emergency function groups complement one another and should be integrated to address the disastrous situation together. During response process of the "11.22" oil pipeline explosion accident, each emergency function group collaborates with others to provide continuous and coordinated emergency services to achieve common incident objectives, such as saving human lives, protecting properties and the environment, stabilizing the incident, restoring basic services and community functions, and establishing a safe and secure environment for transition to recovery. All the participating organizations are classified into 15 groups in this research. Each group provides specific emergency response functions and performs relevant tasks during the emergency response process, as detailed in Table 5. In addition, the affected communities which received emergency service and provide self- and mutual aids to victims are defined as a specific group for conducting structural block analysis. In this research, structural block analysis determines whether interactive relationships exist among these blocks representing emergency function groups.

Figure 5 presents the block model showing interactive relationships within and among blocks representing emergency function groups. Similar to Figure 4, the lines represent the interactive relationships between different emergency function groups or among response organizations within the same group. The width and color of the lines represent the density of network ties, so that a wider and darker line indicates a higher density of ties among or within the emergency function group. From Figure 5, response organizations affiliated with the emergency function groups Resident Resettlement, Psychological Intervention and Mass Care,

TABLE 5: The involved emergency function groups responding to the "11.22" oil pipeline explosion.

Emergency function group	Task description
Traffic Control	Manage the traffic
Transportation	Transport materials to the accident scene
Resident Resettlement	Provide house and food to evacuated residents
Information Issue and Media Management	Issue emergency information to the public and manage the media personnel on scene
Information Monitoring	Collect, analyze, and distribute emergency situation information
Public Works and Engineering	Restore and repair electricity, water, and gas supply systems
Emergency Medical Care	Provide medical care to victims
Psychological Intervention and Mass Care	Provide mental health service to victims
Emergency Command and Coordination	Coordinate incident management and response efforts, issue mission assignments, and plan the emergency response
Firefighting and Search and Rescue	Search for and rescue victims in the accident; extinguish fire caused by the explosion in the accident
Life Material Support	Provide life materials to the evacuated residents
Oil and Hazardous Material Response	Clear the pollution caused by the oil spill on land and in the sea
Guarding and Public Security	Block the scene of the accident and maintain public order
Communication Support	Maintain and restore telecommunication infrastructure
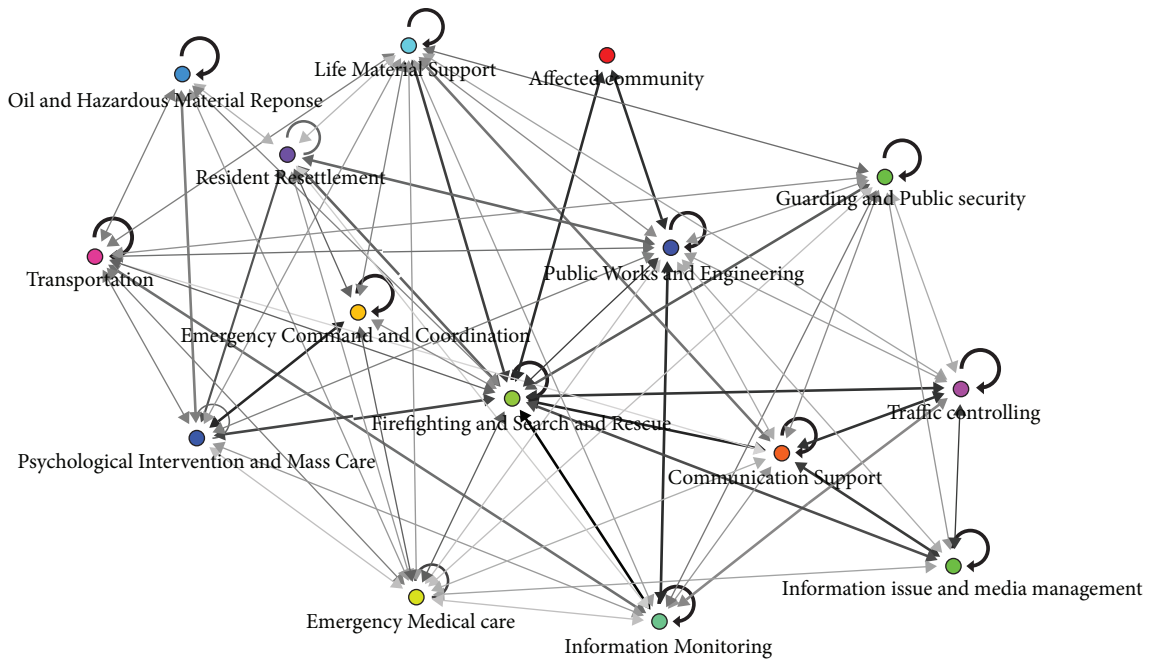
FIGURE 5: Block model of emergency response function groups.

and Emergency Medical Care, interact with each other at a low level, and the interorganizational relationships among organizations affiliated with the same groups should be improved to provide more streamlined emergency response services. As the core of the whole network, there does not exist dense interactive relationships between the block representing the Emergency Command and Coordination group and other groups. That indicates that the former groups cannot coordinate response organizations in the other group to

achieve better network collaboration. Especially, there are no interactive relationships between the Emergency Command and Coordination group and the Affected Community group. This observation indicates that information about the disastrous situation and consequences of this explosion cannot be collected rapidly and effectively during the emergency response. Therefore, that presents barriers to the interorganizational collaboration in the emergency response process.

7. Discussions and Managerial Suggestions

Intergovernmental and interorganizational collaboration taking the form of network ties among response organizations is necessary to successfully address major accidents and determine the performance of emergency response to a great extent. The networked response operations spanning organizational boundaries are a highly complex assembly of multiple heterogeneous organizations with different capabilities and functions for achieving common objectives. Reflecting on the response to the "11.22" oil pipeline explosion accident, responding to a major accident requires that public organizations affiliated multiple government levels, as well as private and nonprofit organizations, collaborate and interact with one another, while still performing tasks within the boundaries of their own organization. Appropriate structure is a necessary condition for achieving network effectiveness. A number of managerial suggestions are drawn from examining constituents and structural characteristics of an actual emergency response network for the "11.22" oil pipeline explosion accident in China but can also be relevant to improve the existing emergency management system in the centralized political-administrative context.

First, in the participating organizations, 31.1% are private organizations, while 26.3% are from the nonprofit sector. Thus, both private and nonprofit organizations play essential roles for addressing major accidents. However, the existing emergency management network at each government level emphasizes the responsibilities and functions of public organizations in emergency response and neglects the participation of the private and nonprofit sectors. The governments should plan ahead to effectively establish and sustain partnerships between other public, private, and nonprofit organizations for integrating their capabilities into the existing emergency management systems at each government level.

Second, from the centrality analysis, public organizations at the county and municipal government levels play the most important roles in addressing major accidents and are located at the central position in the emergency response network. The network analysis results show that governments at the county and municipal government levels do not only rapidly respond to major accidents but also play more important roles than the principal and central governments. Furthermore, from the structural block analysis in Section 6.4.3, the interactive relationships between private and public organizations at the central, provincial, and municipal government levels are weaker than those between private and public sections at the county government level. During response to major accidents, the participating organizations at the local government level are faster than others and are more active than others to coordinate their response operations. Therefore, network leadership capability of local governments should be focused on and enhanced for a successful response to major accidents. The important lesson from this empirical research indicates that more investment should be made at the local government level in the field of emergency management.

Third, despite network mechanism being considered as an alternative to the hierarchy mechanism, this case study reveals that a blend of hierarchical, market, and network principles and strategies is mixed in fostering intergovernmental and interorganizational collaboration during response to major accidents in China. The involved multiple governance mechanisms are embedded in the form of diversified links spanning organizational boundaries as shown in the map of an emergency response network in Section 6.2. These governance mechanisms complement each other for improving the effectiveness of the emergency response network. In particular, the shadow of interorganizational hierarchies is apparent in the network according to the structural block analysis of organizational types in Section 6.4.3. That reveals that the political-administrative system of the Chinese government plays essential roles in the networked response operations for addressing major accidents. In fact, it facilitates rapid collaboration and coordination in these intergovernmental and interorganizational network environments during emergencies.

Finally, participating organizations affiliated with the same emergency function group should interact with each other to provide specific emergency services in this empirical research. Moreover, all the fourteen emergency function groups involved in the responding process are required to keep interaction and collaboration with each other for providing streamlined services to affected communities and achieving better network effectiveness. As discussed in Section 6.4.3, a number of problems are identified from the block analysis of emergency function groups, which provides guidance to improve the design of emergency function groups in the involved emergency operation plans. For example, organizations affiliated with the emergency function groups Resident Resettlement, Psychological Intervention and Mass Care, and Emergency Medical Care should pay more attention to establish and sustain partnerships with each other. The interactive relationships between the Emergency Command and Coordination group and those in other groups are weak and should be improved. In particular, the information reporting relationships from the affected communities to organizations affiliated with the Emergency Command and Coordination group should be established and strengthened for collecting information on disastrous situations and identifying response requirements more exactly and quickly.

8. Conclusions and Future Work

Intergovernmental and multiorganizational collaboration for addressing major accidents involves complex interactions spanning organizational boundaries of public sectors in multiple government levels, as well as private and nonprofit sectors. Despite the institutionalized emergency management systems having been established in the last decade, the lack of interorganizational collaboration and coordination presents challenges to Chinese governments for responding to emergencies involving multiple government levels. In this research, the whole picture of intergovernmental and cross-sectoral collaboration in responding to the well-known oil pipeline explosion accident is focused on and examined from network perspectives to obtain managerial

insights in improving the existing emergency management system in the centralized political-administrative context, such as China.

By conducting mix data collection methods, all the participating organizations and numerous interorganizational relationships in both hierarchical and horizontal dimensions are identified. The emergency response network is built to represent the interorganizational collaboration of varying types involved in the response process. The research facilitates the development of theoretical linkages between the emergency response concepts and social network analysis. From quantitative analysis results of the emergency response network based on SNA, a number of findings and managerial suggestions for improving the existing emergency management system in China are proposed. First, the Chinese government should pay attention to establish and sustain partnerships with private and nonprofit organizations, and a blend of hierarchical, market, and network principles and strategies should be mixed to complement one another in fostering collaboration among responsible organizations in the emergency management system in China. Second, the participating organizations at the local government level are faster in responding to accidents and are more active than other participators in coordinating their response operations. The capabilities of local governments should be emphasized in emergency management. Finally, the interactive relationships among specific emergency function groups and between the possible affected communities and organizations performing the emergency command and coordination function should be strengthened. Although we focus on a case study of a major explosion accident, this research provides insights on how to improve the intergovernmental collaboration involved in addressing complex problems in the centralized political-administrative context.

The main limitations of the reported research pertains to the data sources. Tracking real-world intergovernmental collaboration to identify all the organizations in the network boundaries and to determine interorganizational relationships of varying types is an existing problem. Furthermore, this study focuses on static networks by capturing interactions during the entire response process and disregards the evolution of the emergency response network over time. Actually, the intergovernmental and cross-sectoral response to large-scale emergencies can be reframed as a dynamic and adaptive network that adjusts for best fitting the demands of ever-changing disastrous situations. The future work is to conduct research on the evolution of emergency response networks.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] L. K. Comfort and N. Kapucu, "Inter-organizational coordination in extreme events: the World Trade Center attacks, September 11, 2001," *Natural Hazards*, vol. 39, no. 2, pp. 309–327, 2006.

[2] L. K. Comfort, "Managing intergovernmental responses to terrorism and other extreme events," *Publius*, vol. 32, no. 4, pp. 29–50, 2002.

[3] D. P. Moynihan, "The network governance of crisis response: case studies of incident command systems," *Journal of Public Administration Research and Theory*, vol. 19, no. 4, pp. 895–915, 2009.

[4] N. Kapucu, T. Arslan, and M. L. Collins, "Examining intergovernmental and interorganizational response to catastrophic disasters: toward a network-centered approach," *Administration & Society*, vol. 42, no. 2, pp. 222–247, 2010.

[5] J. M. Brooks, D. Bodeau, and J. Fedorowicz, "Network management in emergency response:articulation practices of state-level managers–interweaving up, down and sideways," *Administration & Society*, vol. 45, no. 8, pp. 911–948, 2012.

[6] A. Boin and F. Bynander, "Explaining success and failure in crisis coordination," *Geografiska Annaler*, vol. 97, no. 1, pp. 123–135, 2016.

[7] N. Kapucu and V. Garayev, "Designing, managing, and sustaining functionally collaborative emergency management networks," *The American Review of Public Administration*, vol. 43, no. 3, pp. 312–330, 2012.

[8] X. Lu and L. Xue, "Managing the unexpected: sense-making in the Chinese emergency management system," *Public Administration*, vol. 94, no. 2, pp. 414–429, 2016.

[9] F. A. Osman, A. M. Shahan, and F. Jahan, "Managing natural disasters in Bangladesh: activating the network approach," *Public Organization Review*, vol. 15, no. 1, pp. 99–116, 2015.

[10] T. E. Drabek, *Emergency Management: Strategies for Maintaining Organizational Integrity*, Springer-Verlag, New York, NY, USA, 1990.

[11] X. Guo and N. Kapucu, "Examining collaborative disaster response in China: network perspectives," *Natural Hazards*, vol. 79, no. 3, pp. 1773–1789, 2015.

[12] K. Jung and M. Song, "Linking emergency management networks to disaster resilience: bonding and bridging strategy in hierarchical or horizontal collaboration networks," *Quality & Quantity*, vol. 49, no. 4, pp. 1465–1483, 2015.

[13] N. Kapucu and F. Demiroz, "Measuring performance for collaborative public management using network analysis methods and tools," *Public Performance & Management Review*, vol. 34, no. 4, pp. 549–579, 2011.

[14] M. P. Mandell and R. Keast, "Evaluating the effectiveness of interorganizational relations through networks: developing a

framework for revised performance measures," *Public Management Review*, vol. 10, no. 6, pp. 715–731, 2008.

[15] K. G. Provan, A. Fish, and J. Sydow, "Interorganizational networks at the network level: a review of the empirical literature on whole networks," *Journal of Management*, vol. 33, no. 3, pp. 479–516, 2007.

[16] N. Kapucu and Q. Hu, "Understanding multiplexity of collaborative emergency management networks," *The American Review of Public Administration*, vol. 46, no. 4, pp. 399–417, 2016.

[17] W. L. Waugh and G. Streib, "Collaboration and leadership for effective emergency management," *Public Administration Review*, vol. 66, no. s1, pp. 131–140, 2006.

[18] O. Noran, "Collaborative disaster management: an interdisciplinary approach," *Computers in Industry*, vol. 65, no. 6, pp. 1032–1040, 2014.

[19] J. M. Bryson, B. C. Crosby, and M. M. Stone, "Designing and implementing cross-sector collaborations: needed and challenging," *Public Administration Review*, vol. 75, no. 5, pp. 647–663, 2015.

[20] C. Ansell and A. Gash, "Collaborative governance in theory and practice," *Journal of Public Administration Research and Theory*, vol. 18, no. 4, pp. 543–571, 2008.

[21] Q. Hu, S. Khosa, and N. Kapucu, "The intellectual structure of empirical network research in public administration," *Journal of Public Administration Research and Theory*, vol. 26, no. 4, pp. 593–612, 2016.

[22] N. Kapucu, "Interorganizational coordination in dynamic context: networks in emergency response management," *Connections*, vol. 50, no. 2, pp. 139–146, 2005.

[23] Web of The Central People's Government of the People's Republic of China, "Act on addressing emergencies of Republic of China," August 2007, http://www.gov.cn/ziliao/flfg/2007-08/30/content_732593.htm.

[24] K. G. Provan and P. Kenis, "Modes of network governance: structure, management, and effectiveness," *Journal of Public Administration Research and Theory*, vol. 18, no. 2, pp. 229–252, 2007.

[25] S. E. Robinson, W. S. Eller, M. Gall, and B. J. Gerber, "The core and periphery of emergency management networks," *Public Management Review*, vol. 15, no. 3, pp. 344–362, 2013.

[26] H. Zhang, X. Zhang, L. Comfort, and M. Chen, "The emergence of an adaptive response network: the April 20, 2013 Lushan, China earthquake," *Safety Science*, vol. 90, pp. 14–23, 2016.

[27] D. F. Kettl, "Managing boundaries in American administration: the collaboration imperative," *Public Administration Review*, vol. 66, no. s1, pp. 10–19, 2006.

[28] C. J. Koliba, R. M. Mills, and A. Zia, "Accountability in governance networks: an assessment of public, private, and nonprofit emergency management practices following Hurricane Katrina," *Public Administration Review*, vol. 71, no. 2, pp. 210–220, 2011.

[29] Q. Hu, C. C. Knox, and N. Kapucu, "What have we learned since September 11, 2001? A network study of the Boston Marathon bombings response," *Public Administration Review*, vol. 74, no. 6, pp. 698–712, 2014.

[30] W. W. Powell, "Neither market nor hierarchy: network forms of organization," *Research in Organizational Behavior*, vol. 12, pp. 295–336, 1990.

[31] R. Yin, *Case Study Research: Design and Methods*, Sage Publication, 2013.

[32] Web of State Administration of work Safety, "Shandong Qingdao Sinopec east yellow "11.22" explosion leakage of oil pipeline especially big accident investigation report," January 2014, http://www.gov.cn/gzdt/2011-12/29/content_2032986.htm.

[33] W. Scott and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, UK, 1994.

[34] Netminerhttp://www.netminer.com/main/main-read.do.

[35] S. E. Robinson, W. Eller, M. Gall, and B. J. Gerber, "The core and periphery of emergency management networks: a multimodal assessment of two evacuation hosting networks from 2000-2009," *SSRN Electronic Journal*, vol. 15, pp. 1–38, 2012.

[36] S. O. Choi and R. S. Brower, "When practice matters more than government plans: a network analysis of local emergency management," *Administration & Society*, vol. 37, no. 6, pp. 651–678, 2006.

[37] K. R. Isett, I. A. Mergel, K. Leroux, P. A. Mischen, and R. K. Rethemeyer, "Networks in public administration scholarship: understanding where we are and where we need to go," *Journal of Public Administration Research and Theory*, vol. 21, Supplement 1, pp. i157–i173, 2011.

[38] K. Chang, "Understanding cross-sector collaboration in emergency management: the dynamics of vertical and horizontal networks," *Dissertations & Theses-Gradworks*, 2012.

Research Article

Exponential Synchronization Control of Discontinuous Nonautonomous Networks and Autonomous Coupled Networks

Chao Yang [ID],[1] Lihong Huang,[2] and Fangmin Li [ID][1,3]

[1]*Department of Mathematics and Computer Science, Changsha University, Changsha 410022, China*
[2]*School of Mathematical and Statistics, Changsha University of Science and Technology, Changsha, Hunan 410114, China*
[3]*School of Information Engineering, Wuhan University of Technology, Wuhan 407003, China*

Correspondence should be addressed to Fangmin Li; lifangmin@whut.edu.cn

This paper concerns complex delayed neural networks with discontinuous activations. Based on the framework of differential inclusion theory, we design two novel controllers by regulating a parameter σ $(0 \leq \sigma < 1)$ which covers both discontinuous and continuous controllers. Then, we investigate a nonautonomous cellular neural network system and autonomous neural network with linear coupling, respectively. By choosing a time-dependent Lyapunov-Krasovskii functional candidate and suitable controllers, some criteria are studied to guarantee the exponential synchronization of the complex delayed dynamical network. Finally, two numerical examples are given to illustrate our theoretical analysis.

1. Introduction

In the past few decades, the dynamical behavior of synchronization phenomena has attracted much attentions because of its potential practical application in general complex networks [1], pattern recognition [2], secure communication [3], combinational optimization [4], biological systems [5], and so on. Up to now, several types of synchronization of complex neural networks have been studied such as asymptotic synchronization [6], finite-time synchronization [7], and exponential synchronization [8–10]. The synchronization phenomena of a complex dynamical network are said to be an important issue in our theoretical analysis and experimental application.

In real world, there are a large number of nodes in the real-world complex networks. Cao et al. in [11, 12] studied the global synchronization of coupled delayed neural networks with constant and hybrid coupling. The authors in [13] designed a coupling term by $D(x_j(t - \tau(t))) - x_i(t - \tau(t))$ and realized the exponential synchronization for complex dynamical networks with sampled data. After that, some literatures are interested in the

synchronization for neural networks with the coupling term $D(x_j(t - \tau(t))) - x_i(t - \tau(t))$; for example, in [14, 15], the authors investigated the synchronization of coupled networks with hybrid coupling, which were composed of constant coupling and a single coupling delay. By this distance, a new unloading method is obtained in global convergence for complete regular coupling configuration. Generally, the coupling structure is designed by a graph which can be unconnected, directed, and undirected.

As we know, many valid control techniques have been extensively applied in the engineering field, such as impulsive control [16], intermittent control [17], feedback control [18], and adaptive control [19]. In recent years, many researchers receive the results on synchronization stabilization of complex chaotic systems and coupled dynamical networks by pinning a suitable control, and most of the existing controllers were designed in the form of $-k\,\mathrm{sign}\,(e(t))$ $|e(t)|^\sigma (0 \leq \sigma < 1)$; we can see that the controller is continuous if $0 < \sigma < 1$ and the controller is discontinuous if $\sigma = 0$, where $e(t)$ is the synchronization error with control strength k. However, few literatures discuss the two types of switching controllers concurrently, and the two categories

are discussed separately or only in the field of Lipschitz conditions. Because there still have been a lot of difficulties in overcoming the exponential synchronization problem when the activation functions are discontinuous but the controllers are not. However, to the best of our knowledge, few papers focus on the synchronization issue of complex networks with nonlinear coupling function, and there are two kinds of controllers such as continuous case and discontinuous case when the activation functions are still discontinuous.

The neural network system of this paper is a general nonautonomous neural network system with discontinuous activations, and we also consider the corresponding autonomous system in this paper. The main contributions are as follows:

(1) In the existing exponential synchronization research, the neuron activation functions were restricted to be continuous and bounded, and the assumptions of the system were complex. So this paper consider a more general neural network model and simpler conditions for gaining the exponential synchronization goal

(2) It is the first time that the exponential synchronization control of the nonautonomous systems with discontinuous activation and the autonomous system with linear coupling function is considered. The algorithm in this paper is optimized, where sufficient conditions formulated by the Lyapunov function are established to gain the exponential synchronization. The theoretical results can also be used in a wider scope

(3) Novel analytical techniques are proposed, and strict mathematical proofs are given for the global exponential synchronization of the discontinuous neural network with coupled and time delays. We design novel discontinuous controllers and continuous controllers in this paper. When both neuron functions and controllers are discontinuous, there is still a lack of complete theory of synchronization

(4) The technique skill and control algorithm are different from those in previous papers (e.g., [20]). We introduce some novel tools such as exponential synchronization theorem, differential inclusion in the sense of Filippov, and generalized Lyapunov approach under a 1-norm framework, and the methods proposed in this paper can be extended to investigate the synchronization of neural network systems

The structure of this paper is outlined as follows. In the next section, we design the model and introduce some basic preliminary lemmas and definitions. In Section 3, we design a continuous controller to realize the exponential synchronization of the nonautonomous network system with discontinuous activations and describe a nonlinear coupling function to guarantee the synchronization issue of the time-delayed discontinuous neural network by considering a discontinuous controller. In Section 4, we give two numerical examples to illustrate our theoretical results. Finally, we conclude this paper in Section 5.

Notation 1. Let $\mathbb{R}^n$ denote the n-dimensional Euclidean space, and let the superscript T denote the transposition. Let $x = (x_1, x_2, \ldots, x_n)^T$ and $y = (y_1, y_2, \ldots, y_n)^T$; by $x > 0$ $(x \geq 0)$, we mean that $x_i > 0 (x_i \geq 0)$ for all $i = 1, 2, \ldots, n$. $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$, $\langle \cdot, \cdot \rangle$ denotes the inner product. If $x \in \mathbb{R}$, $\|x\|$ denotes the vector norm of x, while $\|x\|_1 = \sum_{i=1}^n |x_i|$. Given the real matrix $A = (a_{ij})_{n \times n}$, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ represent the maximal and minimal eigenvalues of A, respectively. Let diag $(\cdots)$ denote the block diagonal matrix, and let sign $(\cdot)$ denote the sign function.

Finally, let $g(t)$ be the continuous function, and we define that

$$
\begin{aligned}
g^{\max} &= \sup_{t \in \mathbb{R}} |g(t)|, \\
g^{\min} &= \inf_{t \in \mathbb{R}} |g(t)|.
\end{aligned}
\tag{1}
$$

2. Preliminaries

In this section, we give some definitions and preliminary lemmas. The main references are the framework of Filippov, set valued maps, differential inclusion, and so on [21–26]. Firstly, we consider the discontinuous function f to introduce the solution of the system, and we denote the closure of the convex hull of X as $K[X]$; we can expand the property of the Filippov solution to the system.

By the discussions in Section 1, in this paper, we consider the following general nonautonomous neural network system with time-varying delays and discontinuous right-hand sides:

$$
\begin{aligned}
\frac{dx_i(t)}{dt} = &-a_i(t)x_i(t) + \sum_{j=1}^n b_{ij}(t)f_j(x_j(t)) \\
&+ \sum_{j=1}^n c_{ij}(t)f_j(x_j(t - \tau_{ij}(t))) + I_i(t), \quad i = 1, 2, \ldots, n,
\end{aligned}
\tag{2}
$$

where $x_i(t)$ corresponds to the state vector of the ith unit at time t, $a_i(t) > 0$ denotes the self-inhibition with which the ith neuron will reset its potential to the resting state in isolations when disconnected from the network and inputs, $b_{ij}(t)$ and $c_{ij}(t)$ represent the connection strength and the delayed connection strength of the jth neuron on the ith neuron, respectively, $f_j(x_j(t))$ represents the activation function and the time-delayed activation function of jth neuron, $I_i(t)$ is a constant external input vector, $\tau_{ij}(t)$ corresponds to the transmission delay of the ith unit along the axon of the jth unit at time t and is a continuously differentiable function, and there exist $\tau =$

$\max_{1 \le i,j \le n} \{\max_{t \in [0,\omega]} \tau ij(t)\} \ge 0$ and a negative constant τ_{ij}^D satisfying

$$0 \le \tau_{ij}(t) \le \tau,$$
$$\dot{\tau}_{ij}(t) \le \tau_{ij}^D < 1. \tag{3}$$

Moreover, we obtain an autonomous system when coefficients are reduced to constants corresponding to model (2) as follows:

$$\frac{dx_i(t)}{dt} = -a_i x_i(t) + \sum_{j=1}^{n} b_{ij} f_j(x(t))$$
$$+ \sum_{j=1}^{n} c_{ij} f_j(x_j(t - \tau_{ij}(t))) + I_i, \quad i = 1, 2, \dots, n. \tag{4}$$

Equivalently, the differential equation system can be transformed into the following matrix format:

$$\frac{dx(t)}{dt} = -Ax(t) + Bf(x(t)) + Cf(x(t - \tau(t))) + I, \tag{5}$$

where $A = \text{diag}(a_1, a_2, \dots, a_n)$, $B = (b_{ij})_{n \times n}$, and $C = (c_{ij})_{n \times n}$.

To establish our main results, we assume the following basic conditions for the neuron activations in model (2) or (4):

Assumption 1. For every $j = 1, 2, \dots, n$, f_j is continuous except for a countable set of isolate jump discontinuous points ρ_k, where there exist finite right and left limits, and in every compact set of R, it has only a finite number of jump discontinuous points.

Definition 1. A vector function $x = (x_1, x_2, \dots, x_n)^T : [-\tau, T) \to \mathbb{R}^n$, $T \in (0, +\infty]$, is a state solution of the discontinuous system (2) on $[-\tau, T)$ if

(1) x is continuous on $[-\tau, T)$ and absolutely continuous on any compact interval of $[0, T)$

(2) there exists a measurable function $\gamma_j(t) \in K[f_j(x(t))]$ for a.e. $t \in [-\tau, T)$ and

$$\frac{dx_i(t)}{dt} = -a_i(t) x_i(t) + \sum_{j=1}^{n} b_{ij}(t) \gamma_j(t)$$
$$+ \sum_{j=1}^{n} c_{ij}(t) \gamma_j(t - \tau_{ij}(t)) + I_i(t), \quad t \in [0, T). \tag{6}$$

Any function $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)^T$ satisfying (6) is called an output solution associated with the state $x = (x_1, x_2, \dots, x_n)^T$; then, in the sense of Filippov, we point

out that the state x is a solution of (2) for a.e. $t \in [0, T)$ and we obtain the following differential inclusion:

$$\frac{dx_i(t)}{dt} \in -a_i(t) x_i(t) + \sum_{j=1}^{n} b_{ij}(t) K[f_j(x_j(t))]$$
$$+ \sum_{j=1}^{n} c_{ij}(t) K[f_j(x_j(t - \tau_{ij}(t)))] + I_i(t), t \in [0, T). \tag{7}$$

Definition 2. The network is said to achieve global exponential synchronization if there exist some constants $\lambda > 0$, $T > 0$, and $M_0 > 0$ such that for any initial values $\phi_i(s)$ $(i = 1, 2, \dots, n)$,

$$\|x_j(t) - x_i(t)\| \le M_0 e^{-\lambda t} \tag{8}$$

hold for all $t > T$ and for any $i, j = 1, 2, \dots, n$.

Lemma 1 (see [10]). *If $V(x): \mathbb{R}^n \to \mathbb{R}$ is C-regular and $x(t): [0, +\infty) \to \mathbb{R}^n$ is absolutely continuous on any compact subinterval of $[0, +\infty)$. Then, $x(t)$ and $V(x(t)): [0, +\infty) \to \mathbb{R}$ are differentiable for almost all $t \in [0, +\infty)$ and*

$$\frac{dV(x(t))}{dt} = \left\langle \varsigma(t) \frac{dx(t)}{dt} \right\rangle, \quad \forall \varsigma(t) \in \partial V(x(t)). \tag{9}$$

Lemma 2 (see [11, 12]). *Given an undirected graph F with the adjacency matrix $C = [c_{ij}]$ and Laplacian matrix L, equality*

$$x^T L x = \frac{1}{2} \sum_{i,j=1}^{n} c_{ij}(x_i - x_j)^2 \tag{10}$$

holds for arbitrary $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

Let $\mathbb{F}(x) \triangleq K[f(x)] = (K[f_1(x)], K[f_2(x)], \dots, K[f_n(x)])$, where $K[f_i(x)] = [\min\{f_i(x^-), f_i(x^+)\}, \max\{f_i(x^-), f_i(x^+)\}]$. Then, we assume the neuron activation functions in (2) or (4) to satisfy the following condition:

Assumption 2. For $x, y \in \mathbb{R}$, there exist nonnegative constants α and β such that

$$\|\mathbb{F}[f(x) - f(y)]\| = \sup_{\xi \in \mathbb{F}[f(x) - f(y)]} \|\xi\| \le a\|x - y\| + \beta. \tag{11}$$

3. Main Results

In this section, the discontinuous controller and continuous controller are designed; then, we divide this section into two parts to derive the global exponential synchronization conditions of discontinuous nonautonomous networks and autonomous coupled networks, respectively.

3.1. Exponential Synchronization with the Continuous Controller. Firstly, we consider the nonautonomous neural

network model (6) as the driver system, and the controlled response system can be described as follows:

$$\frac{dy_i(t)}{dt} = -a_i(t)y_i(t) + \sum_{j=1}^{n} b_{ij}(t)f_j\left(y_j(t)\right)$$
$$+ \sum_{j=1}^{n} c_{ij}(t)f_j\left(y_i\left(t - \tau_{ij}(t)\right)\right) + I_i(t) + u_i(t), \quad i = 1, 2, \ldots, n,$$

$$(12)$$

where $u_i(t)$ is the controller to be designed for realizing the synchronization of the driver response system. The other parameters are the same as those in model (6).

Our first goal is to drive the response network model (12) to synchronize with the nonautonomous network model (6) with continuous controllers. To this end, choose the parameter $0 < \sigma < 1$, and the continuous controller $u_i(t)$ is given by

$$u_i(t) = -k_1(y_i(t) - x_i(t)) - k_2 \operatorname{sign}(y_i(t) - x_i(t))|y_i(t) - x_i(t)|^\sigma.$$

$$(13)$$

Then, by subtracting (6) from (12), let $e_i(t) = y_i(t) - x_i(t)$. In view of Assumption 1 and Definition 1, by differential inclusions and set valued maps, we can see that there exists a measurable function $\xi_j(t) \in K[f_j(y_j(t))]$ for a.e. $t \in [0, T]$ and we can obtain the synchronization error system as follows:

$$\frac{de_i(t)}{dt} = -a_i(t)e_i(t) + \sum_{j=1}^{n} b_{ij}(t)\Gamma_j(t)$$
$$+ \sum_{j=1}^{n} c_{ij}(t)\Gamma_j\left(t - \tau_{ij}(t)\right)) - k_1 e_i(t) - k_2 \operatorname{sign}(e_i(t))|e_i(t)|^\sigma,$$

$$(14)$$

where $\Gamma_j(t) = \xi_j(t) - \gamma_j(t)$ and $\Gamma_j(t - \tau_{ij}(t)) = \xi_j(t - \tau_{ij}(t)) - \gamma_j(t - \tau_{ij}(t))$.

Then, we give the following theorem to derive the response network system (6) with $0 < \sigma < 1$ synchronizing with the driver network system (2). Before doing this, we give a further condition on the discontinuous activation function f_j as follows:

Theorem 1. *If Assumptions 1 and 2 hold, the nonautonomous discontinuous neural networks achieve global exponential synchronization under the continuous switching controller (13) with $0 < \sigma < 1$; if there exist positive $\zeta_1, \zeta_2, \ldots, \zeta_n$ and a very small positive constant $\varepsilon > 0$, for $i = 1, 2, \ldots, n$, the following conditions are satisfied:*

$$\lim_{t \to +\infty} \sup Q_i(t) < 0, \tag{15}$$

where

$$Q_i(t) = \zeta_i b_{ii}(t) + \sum_{j=1, j\neq i}^{n} \zeta_j |b_{ij}(t)| + \sum_{j=1}^{n} \zeta_j e^{\varepsilon\tau} \frac{\left|c_{ij}\left(\varphi_{ij}^{-1}(t)\right)\right|}{1 - \dot{\tau}_{ij}\left(\varphi_{ij}^{-1}(t)\right)}.$$

$$(16)$$

Proof 1. Consider the following candidate Lyapunov function:

$$V(t) = e^{\varepsilon t} \sum_{i=1}^{n} \varsigma_i |e_i(t)| + \sum_{i,j=1}^{n} \varsigma_i$$
$$\times \int_{t-\tau_{ij}(t)}^{t} \frac{\left|c_{ij}\left(\varphi_{ij}^{-1}(s)\right)\right|}{1 - \dot{\tau}_{ij}\left(\varphi_{ij}^{-1}(s)\right)} \left|\Gamma_j(s)e^{\varepsilon(s+\tau)}\right| ds,$$

$$(17)$$

where φ_{ij}^{-1} is the inverse function of $\varphi_{ij}(t) = t - \tau_{ij}(t)$.

Note that the function $|e_i(t)|$ is locally continuous (Lipschitz) in e_i on R; then, we can see that $V(e(t))$ is regular. According to the definition of Clarke's generalized gradient of the absolute value function $|e_i(t)|$ at $e_i(t)$, we obtain that there exist $\partial(|e_i(t)|) = K[\operatorname{sign}(e_i(t))] = 1$ if $e_i(t) < 0$, $\partial (|e_i(t)|) = K[\operatorname{sign}(e_i(t))] = -1$ if $e_i(t) > 0$, and $\partial(|e_i(t)|) = K[\operatorname{sign}(e_i(t))] = [-1, 1]$ if $e_i(t) = 0$. For any $\vartheta_i(t) \in K[\operatorname{sign}(e_i(t))]$, we have $\vartheta_i(t) = \operatorname{sign}(e_i(t))$, if $e_i(t) \neq 0$; $\vartheta_i(t)$ can arbitrarily be selected in $[-1, 1]$, if $e_i(t) = 0$.

Then, by Lemma 1 and calculating the time derivative of $V(t)$, we obtain that

$$\frac{dV(t)}{dt} = \varepsilon e^{\varepsilon t} \sum_{i=1}^{n} \varsigma_i |e_i(t)| + e^{\varepsilon t} \sum_{i=1}^{n} \varsigma_i \operatorname{sign}(e_i(t)) \cdot$$
$$\cdot \left\{ -a_i(t)e_i(t) + \sum_{j=1}^{n} b_{ij}(t)\Gamma_j(t) \right.$$
$$+ \sum_{j=1}^{n} c_{ij}(t)\Gamma_j\left(t - \tau_{ij}(t)\right) - k_1 |e_i(t)|$$
$$\left. - k_2 \operatorname{sign}(e_i(t))|e_i(t)|^\sigma \right\}$$
$$+ \sum_{i,j=1}^{n} \varsigma_i \frac{\left|c_{ij}\left(\varphi_{ij}^{-1}(t)\right)\right|}{1 - \dot{\tau}_{ij}\left(\varphi_{ij}^{-1}(t)\right)} \left|\Gamma_j(t)e^{\varepsilon(t+\tau)}\right|$$
$$- \sum_{i,j=1}^{n} \varsigma_i |c_{ij}(t)| |\Gamma_j\left(t - \tau_{ij}(t)\right)| \left|e^{\varepsilon\left(t+\tau-\tau_{ij}(t)\right)}\right|$$
$$\leq - \sum_{i=1}^{n} \varsigma_i e^{\varepsilon t}(k_1 + a_i(t) - \varepsilon)|e_i(t)| + \sum_{i=1}^{n} \varsigma_i e^{\varepsilon t} b_{ii}(t) |\Gamma_j(t)|$$
$$+ \sum_{i=1}^{n} \sum_{j\neq i}^{n} \varsigma_i e^{\varepsilon t} |b_{ij}(t)| |\Gamma_j(t)|$$

$$+ \sum_{i,j=1}^{n} \varsigma_i e^{\varepsilon(t+\tau)} \frac{\left| c_{ij}\left(\varphi_{ij}^{-1}(t)\right) \right|}{1 - \dot{\tau}_{ij}\left(\varphi_{ij}^{-1}(t)\right)} \left| \Gamma_j(t) \right| - k_2 |e_i(t)|^{\sigma}$$

$$= -\sum_{i=1}^{n} \varsigma_i e^{\varepsilon t}\left(k_1 + a_i(t) - \varepsilon\right)|e_i(t)|$$

$$+ e^{\varepsilon t} \sum_{i=1}^{n} \left\{ \varsigma_i b_{ii}(t) + \sum_{j=1, j\neq i}^{n} \varsigma_i |b_{ij}(t)| \right.$$

$$+ \sum_{j=1}^{n} \varsigma_j e^{\varepsilon t} \frac{\left| c_{ij}\left(\varphi_{ij}^{-1}(t)\right) \right|}{1 - \dot{\tau}_{ij}\left(\varphi_{ij}^{-1}(t)\right)} \Bigg\} \left| \Gamma_j(t) \right| - k_2 |e_i(t)|^{\sigma}$$

$$\leq -\sum_{i=1}^{n} \varsigma_i e^{\varepsilon t}\left(k_1 + a_i^{\min} - \varepsilon\right)|e_i(t)| + e^{\varepsilon t} \sum_{i=1}^{n} Q_i(t)|\Gamma_i(t)|$$

$$- k_2 |e_i(t)|^{\sigma}.$$

$$(18)$$

By the assumption of the theorem, ε can be a very small positive constant, and we can see that there exist positive constants θ_i and $t_0 \geq 0$ such that if $t \geq t_0$,

$$Q_i(t) \leq \theta_i \leq 0. \tag{19}$$

Then, let $\theta_0 = \min\{-\theta_1, -\theta_2, \ldots, -\theta_n\}$, and we deduce that

$$\dot{V}(t) \leq -\min_{1\leq i \leq n} \left\{ \varsigma_i \left(k_i + a_i^{\min} - \varepsilon\right) \right\} e^{\varepsilon t} \sum_{i=1}^{n} |e_i(t)|$$

$$- \theta_0 e^{\varepsilon t} \sum_{i=1}^{n} |\Gamma_i(t)| - k_2 |e_i(t)|^{\sigma} \leq 0, \tag{20}$$

which implies that

$$\sum_{i=1}^{n} |e_i(t)| \leq \frac{V(e_0, 0)}{\min\{\zeta_1, \zeta_2, \ldots, \zeta_n\}} e^{-\varepsilon t}. \tag{21}$$

By Definition 2, the synchronization error $e(t)$ converges to zero. That is to say, the nonautonomous discontinuous and delayed neural networks (2) and (4) can achieve the global exponential synchronization under the continuous switching controller (13). The proof is completed.

Remark 1. Unlike the previous studies, a great difference in our model is that we permit the neuron activation to be discontinuous and unbounded. One can see that the nonlinear function f in this paper may not satisfy the Lipschitz condition any more. There are few results on the synchronization problem if the activations are discontinuous and the controllers are continuous. Our studies extend the previous researches.

3.2. Exponential Synchronization with the Discontinuous Controller. In this part, we describe the following

corresponding N-coupled time-delayed neural networks of (5):

$$\frac{dz_i(t)}{dt} = -Az_i(t) + Bf(z_i(t)) + Cf(z_i(t-\tau)) + I(t)$$

$$+ m \sum_{j=1}^{N} d_{ij}\Phi\varphi(z_j - z_i), \tag{22}$$

where $z_i(t) = (z_{i1}(t), z_{i2}(t), \ldots, z_{in}(t))^T \in \mathbb{R}^n$ ($i = 1, 2, \ldots, N$) denotes the state variable of the ith neuron at time t, m is the coupling strength, $\Phi = \text{diag}(\phi_1, \phi_2, \ldots, \phi_n)$ with $\phi_l > 0$, $l = 1, 2, \ldots, n$, $\varphi(s)$ is the coupling function, $D = [d_{ij}]$ denotes the adjacency matrix of subsystems, where the corresponding Laplacian matrix is represented as L, and all of them are applicable to undirected weighted networks.

Moreover, in order to realize exponential synchronization, a suitable coupling function is important to improve the network performance. Our goal is to derive the coupled time-delayed neural networks with discontinuous controllers synchronizing with the isolated neural network (5). To this end, in this paper, we consider the following coupled neural networks:

$$\frac{dz_i(t)}{dt} = -Az_i(t) + Bf(z_i(t)) + Cf(z_i(t-\tau)) + I(t)$$

$$+ m \sum_{j=1}^{N} d_{ij}\Phi\varphi(z_i - z_i) + v_i(t), \tag{23}$$

where $D = (d_{ij})_{N\times N} \in \mathbb{R}^{N\times N}$ with $d_{ij} > 0$ ($i \neq j$) and $d_{ij} = 0$ ($i, j = 1, 2, \ldots, N$) and $v_i(t)$ is the control algorithm vector similar to (13) when $\sigma = 0$ for the strongly connected network topology which is given as follows:

$$v_i(t) = -k_1(z_i(t) - x(t)) - k_2 \,\text{sign}\,(z_i(t) - x(t)), \tag{24}$$

where k_1 and k_2 are the gain coefficients to be determined. We can see that the controller $v_i(t)$ is discontinuous when $\sigma = 0$.

Then, we choose the discontinuous controller with $\sigma = 0$, and we define the linear coupling function $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ as

$$\varphi(s) = s. \tag{25}$$

Then, the coupled time-delayed complex network can be described as follows:

$$\frac{dz_i(t)}{dt} = -Az_i(t) + Bf(z_i(t)) + Cf(z_i(t-\tau)) + I(t)$$

$$+ m \sum_{j=1, j\neq i}^{N} d_{ij}\Phi z_j(t) + v_i(t), \tag{26}$$

where $\Phi = \text{diag}(\phi_1, \phi_2, \ldots, \phi_n)$ with $\phi_l > 0, l = 1, 2, \ldots, n$.

Similarly, let $w_i(t) = z_i(t) - x(t)$, and we choose the novel discontinuous switching controller (24) and the linear function (25). Also, by differential inclusions and set valued maps,

when $i = 1, 2, \ldots, N$, we can obtain the error dynamical system as follows:

$$\frac{dw_i(t)}{dt} = -Aw_i(t) + B\tilde{\gamma}_i(t) + C\tilde{\gamma}_i(t - \tau))$$
$$+ m \sum_{j=1,j\neq i}^{N} d_{ij}\Phi w_j(t) - k_i w_i(t) - k_2 \text{SIGN}(w_i(t)),$$
(27)

where $\text{SIGN}(w_i(t)) = (\text{SIGN}(w_i(t)), \text{SIGN}(w_{i2}(t)), \ldots, \text{SIGN}(w_{\text{in}}(t)))^T$ with $\text{SIGN}(s) = -1$ if $s < 0$, $\text{SIGN}(s) = [-1, 1]$ if $s = 0$, and $\text{SIGN}(s) = 1$ if $s > 0$ and $\tilde{\gamma}_i(t) = (\tilde{\gamma}_{i1}(t), \tilde{\gamma}_{i2}(t), \ldots, \tilde{\gamma}_{\text{in}}(t))^T = (\xi_{i1}(t) - \gamma_{i1}(t), \xi_{i2}(t) - \gamma_{i2}(t), \ldots, \xi_{\text{in}}(t) - \gamma_{\text{in}}(t))^T$.

Theorem 2. *If Assumptions 1 and 2 hold, we give the further condition:*

Assumption 3. $\min/(1 \le k \le n)\{k_1 + a_k - \sum_{l=1}^{n} a|b_{kl}| - \sum_{l=1}^{n} a|c_{kl}|\} > 0$ and $\min/(1 \le k \le n)\{k_2 - \sum_{l=1}^{n} \beta|b_{kl}| - \sum_{l=1}^{n} \beta|c_{kl}|\} > 0$.

Then, by choosing the coupling function (12), the coupled networks (26), and the isolated model (5), the exponential synchronization under the discontinuous controller (24) with $\sigma = 0$ can be realized.

Proof 2. Define a candidate Lyapunov function as follows:

$$V(t) = V(w(t)) = e^{\varepsilon t} \sum_{i=1}^{N} \|w_i(t)\|_1$$
$$+ \sum_{i=1}^{N} \sum_{k,l=1}^{n} \int_{t-\tau}^{t} e^{\varepsilon(s+\tau)} |c_{kl}||\tilde{\gamma}_{il}(s)| ds,$$
(28)

where $\|w_i(t)\|_1 = \sum_{k=1}^{n} |w_{ik}(t)|$. Similar to Proof 1, we denote

$$\frac{dV(t)}{dt} = V(e(t)) = \varepsilon e^{\varepsilon t} \sum_{i=1}^{N} \sum_{k=1}^{n} |w_{ik}(t)|$$
$$+ e^{\varepsilon t} \sum_{i=1}^{N} \sum_{k=1}^{n} \frac{dw_{ik}(t)}{dt} \cdot \vartheta_{ik}(t)$$
$$+ \sum_{i=1}^{N} \sum_{k,l=1}^{n} e^{\varepsilon(t+\tau)} |c_{kl}||\tilde{\gamma}_{il}(t)| - \sum_{i=l}^{N} \sum_{k,l=1}^{n} e^{\varepsilon t} |c_{kl}||\tilde{\gamma}_{il}(t - \tau)|$$
$$= \varepsilon e^{\varepsilon t} \sum_{i=1}^{N} \sum_{k=1}^{n} |w_{ik}(t)| + e^{\varepsilon t} \sum_{i=1}^{N} \sum_{k=1}^{n} \text{sign}(w_{ik}(t)) \cdot$$
$$\cdot \left\{ -a_k w_{ik}(t) + \sum_{l=1}^{n} b_{kl} \tilde{\gamma}_{il}(t) + \sum_{l=1}^{n} c_{kl} \tilde{\gamma}_{il}(t - \tau) \right.$$
$$\left. + m \sum_{j=1,j\neq i}^{N} d_{ij}\phi_k w_{jk}(t) - k_1 w_{ik}(t) - k_2 \text{sign}(w_{ik}(t)) \right\}$$

$$+ \sum_{i=1}^{N} \sum_{k,l=1}^{n} e^{\varepsilon(t+\tau)} |c_{kl}||\tilde{\gamma}_{il}(t)| - \sum_{i=l}^{N} \sum_{k,l=1}^{n} e^{\varepsilon t} |c_{kl}||\tilde{\gamma}_{il}(t - \tau)|$$
$$\le \varepsilon e^{\varepsilon t} \sum_{i=1}^{N} \sum_{k=1}^{n} |w_{ik}(t)| + e^{\varepsilon t} \sum_{i=1}^{N} \sum_{k=1}^{n} (-a_k |w_{ik}(t)|$$
$$+ \sum_{l=1}^{n} |b_{kl}||\tilde{\gamma}_{il}(t)|| \text{sign}(w_{ik}(t))| + \sum_{l=1}^{n} e^{\varepsilon \tau} |c_{kl}||\tilde{\gamma}_{il}(t)|)$$
$$+ m \sum_{j=1,j\neq 1}^{N} d_{ij}\phi_k |w_{jk}(t)| - k_1 |w_{ik}(t)|$$
$$- k_2 |\text{sign}(w_{ik}(t))| \le -e^{\varepsilon t} \sum_{i=l}^{N} \sum_{k=1}^{n}$$
$$\cdot \left(k_1 + a_k - \varepsilon \quad - \sum_{l=1}^{n} a|b_{kl}| - \sum_{l=1}^{n} e^{\varepsilon t} a|c_{kl}| \right) |w_{ik}(t)|$$
$$+ m \sum_{j=1,j\neq i}^{N} d_{ij}\phi_k |w_{jk}(t)| - e^{\varepsilon t} \sum_{i=l}^{N} \sum_{k=1}^{n}$$
$$\cdot \left(k_2 - \sum_{l=1}^{n} \beta|b_{kl}| - \sum_{l=1}^{n} \beta e^{\varepsilon t} |c_{kl}| \right) |\text{sign}(w_{ik}(t))|.$$
(29)

By Lemma 2 and the property of adjacency matrix D, we deduce that

$$m \sum_{i=l}^{N} \sum_{j=1}^{N} d_{ij}\Phi w_j(t) \le m \sum_{k=1}^{n} \phi k \left[\sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} |w_{jk}(t)| \right]$$
$$= -m \sum_{k=1}^{n} \phi_k \sum_{i=l}^{N} \sum_{j=1,j\neq i}^{N} d_{ij} \left| w_{jk}^T - w_{jk}^T \right| \le 0.$$
(30)

Then, from (30), we deduce that

$$\frac{dV(t)}{dt} \le -e^{\varepsilon t} \sum_{i=1}^{N} \chi_1 |w_{ik}(t)| - e^{\varepsilon t} \sum_{i=l}^{N} \chi_2 |\text{sign}(w_{ik}(t))|, \quad (31)$$

where $\chi_1 = \min_{1<k<n}\{k_1 + a_k - \varepsilon - \sum_{l=1}^{n} a|b_{kl}| - \sum_{l=1}^{n} e^{\varepsilon \tau} a|c_{kl}|\}$ and $\chi_2 = \min_{1<k<n}\{k_2 - \sum_{l=1}^{n} \beta|b_{kl}| - \sum_{l=1}^{n} \beta e^{\varepsilon \tau} |c_{kl}|\}$. By the assumption of the theorem, there must exist a small enough positive $l = 1$ constant ε, such that $\chi_1 > 0$ and $\chi_2 > 0$, which implies

$$\frac{dV(t)}{dt} \le 0, \quad \text{for a.e. } t \ge 0, \quad (32)$$

which yields $V(w(t)) \le V(w(0))$, meaning that $V(w(t))$ is bounded; then, we have

$$\sum_{i=1}^{N} \|w_i(t)\|_1 \le V(w_0, 0) e^{-\varepsilon t}. \quad (33)$$

By Definition 2, the synchronization error $w(t)$ converges to zero. That is to say, the coupled discontinuous and delayed neural networks (26) can be globally exponentially synchronized with the isolated model (5) under the discontinuous switching controller (24). The proof is completed.

Remark 2. In Proof 2, we choose the linear coupling function $\varphi(s) = s$, without the loss of generality, even if the coupling function becomes more complex such as nonlinear function or coupling delay function; many synchronization criteria for delay dependence were derived under these circumstances [20, 27, 28]. In the existing literatures, when the neuron functions were discontinuous, the only thing discussed is a single case for either $\sigma = 0$ or $0 < \sigma < 1$, respectively. When both neuron functions and controllers are discontinuous, there is still no complete conclusion of the issue of synchronization. In this paper, we discuss the exponential synchronization problem of the time-delayed neural network with discontinuous activations under a unified framework of $0 \leq \sigma < 1$.

4. Examples and Simulation Experiment

In this section, to show the effectiveness of our proposed method, two numerical examples are introduced to demonstrate its validity.

Example 1. We consider the following 2-dimensional nonautonomous complex network system:

$$
\begin{cases}
\dfrac{dx_1(t)}{dt} = -x_i(t) - (3 + \cos t)f(x_1(t)) \\
\qquad + \left(\dfrac{1}{4} + \dfrac{1}{4}\cos t\right)f(x_2(t)) \\
\qquad + \left(\dfrac{1}{3} + \dfrac{1}{6}\sin t\right)f(x_1(t - \tau_{11}(t))) \\
\qquad + \left(\dfrac{1}{2} + \dfrac{1}{2}\sin t\right)f(x_2(t - \tau_{12}(t))) + 4, \\
\dfrac{dx_2 t}{dt} = -x_2(t) + \cos t\, f(x_1(t)) - (3 + \sin)f(x_2(t)) \\
\qquad + \dfrac{1}{2}\sin t\, f(x_1(t - \tau_{21}(t))) + 3 + \cos t.
\end{cases}
$$
(34)

Therefore, we can see that $a_1^L = a_2^L = 1$, $b_{11}^M = b_{22}^M = -2$, $c_{11}^M = c_{21}^M = 1/2$, $b_{21}^M = c_{12}^M = 1$, $b_{12}^M = 1/2$, and $c_{22}^M = 0$. The discontinuous activation function can be described as $f(s) = s + \text{sign}(s)$. Let $\tau_{ij}(t) = 1\ (i, j = 1, 2)$. We choose the switching continuous controller $u_i(t) = -e_i(t) - \text{sign}(e_i(t))|e_i(t)|^{1/2}$. Then, Figure 1 shows the time evolution of variables $x_1(t)$ and $x_2(t)$ for the driver neural networks (34); moreover, we can see that the exponential synchronization between the driver system (34) and the corresponding response system can be achieved in Figure 1, which is suitable for our results.

Example 2. We consider three-dimensional autonomous coupled complex dynamical networks as follows:

$$
\begin{cases}
\dfrac{dx_1(t)}{dt} = -x_1(t) - \dfrac{1}{2}f(x_1(t)) + f(x_2(t)) - \dfrac{1}{10}f(x_1(t-1)) \\
\qquad + \dfrac{1}{4}f(x_3(t-1)), \\
\dfrac{dx_2(t)}{dt} = -x_2(t) + \dfrac{1}{3}f(x_2(t)) - \dfrac{1}{5}f(x_3(t)) + \dfrac{1}{4}f(x_2(t-1)), \\
\dfrac{dx_3(t)}{dt} = -x_3(t) + \dfrac{1}{5}f(x_1(t)) - \dfrac{1}{8}f(x_2(t)) + \dfrac{1}{2}f(x_3(t)) \\
\qquad + \dfrac{1}{6}f(x_2(t-1)) + \dfrac{1}{4}f(x_3(t-1)).
\end{cases}
$$
(35)

The discontinuous activation functions are taken as

$$
f(s)\begin{cases} 0.1s - 0.5, & s \geq 0, \\ 0.1s + 0.5, & s < 0. \end{cases}
$$
(36)

Then, let $\alpha = 0.1$ and $\beta = 0.5$, and it is obvious that the conditions (Assumptions 1 and 2) are satisfied. Let the coupling strength be $m = 1$; we choose random switching rules for the coupled networks, and their topologies are illustrated as follows:

(37)
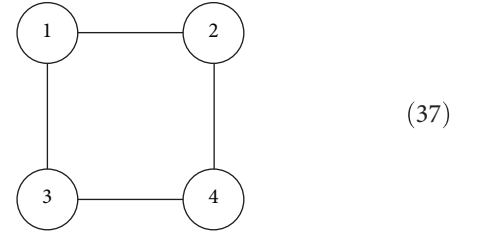
where the adjacency matrix D is easily denoted as

$$
D = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}.
$$
(38)

Then, we consider the discontinuous controller $v_i = -e_i(t) - 2\,\text{sign}(e_i(t))$ with $2k_1 = k_2 = 2$; by substituting the above parameters, we can see that the condition (Assumption 3) holds. We can see that the exponential synchronization between the driver system (35) and the corresponding response system can be depicted in Figure 2, which is suitable for our results.

5. Conclusions

In this paper, we investigate the exponential synchronization of a class of complex dynamical networks based on the framework of nonsmooth analysis and novel technique analysis. By adding a continuous switching controller, we

(a)

(b)

(c)

(d)
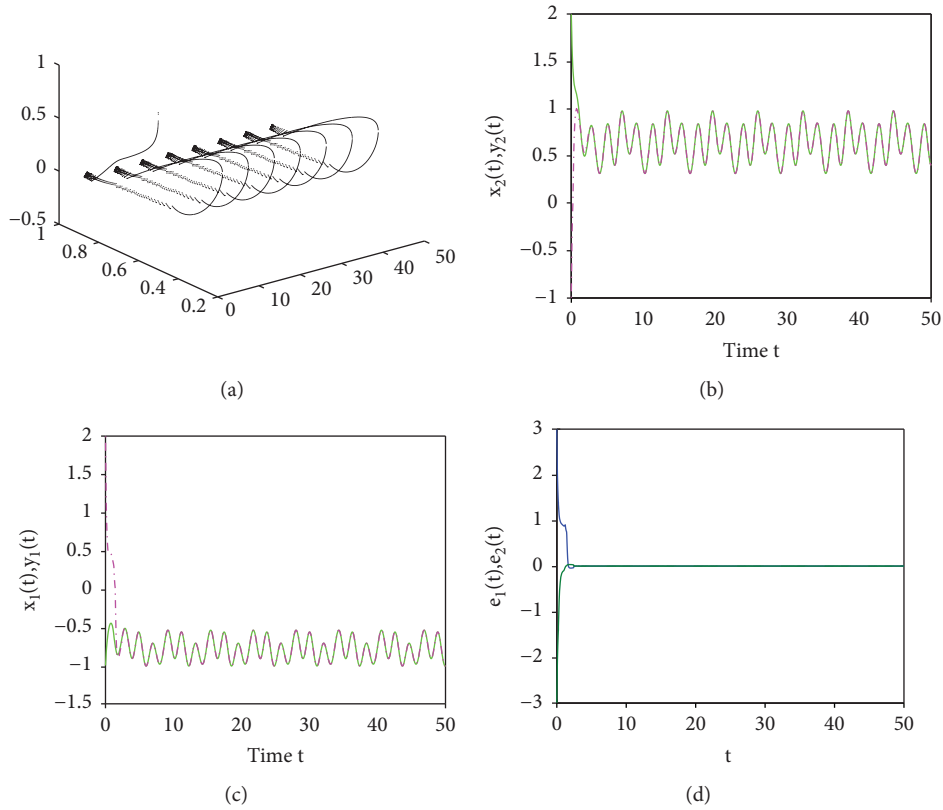
FIGURE 1: (a) The three-dimensional trajectory of state variables x_1 and x_2. (b–c) The time evolution for the driver network system and corresponding response system (34). (d) The time response of the synchronization error between the driver system (34) and corresponding response system with the continuous controller.

(a)

(b)

(c)

(d)
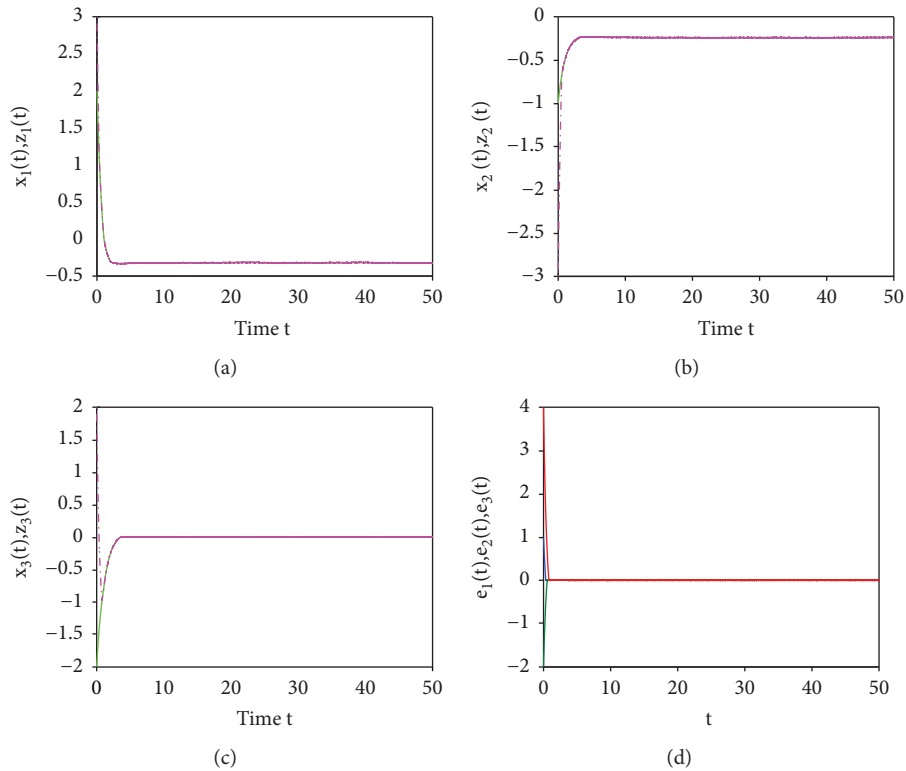
FIGURE 2: (a–c) The time evolution for the driver network system (35) and corresponding response system. (d) The time response of the synchronization error between the driver system (35) and corresponding response system with the discontinuous controller.

realize the global exponential synchronization of the nonautonomous discontinuous and delayed neural networks. Then, we choose a linear coupling function, and the autonomous complex dynamical network can be globally exponentially synchronized with the isolated model under the discontinuous switching controller, by constructing a C-regular Lyapunov-like function which is time-dependent. However, it is not easy to go beyond the conventional Lyapunov function for achieving the exponential synchronization goal. This paper overcomes the limitation of traditional controllers and proposes some novel discontinuous controllers. Moreover, the results have been verified by the numerical examples and computer simulations. In short, our results are provided with an important application significance in the design of synchronized complex dynamical networks.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

We declare that there is no conflict of interest regarding the publication of this paper. And data sharing allows researchers to verify the results of the article, replicate the analysis, and conduct secondary analyses.

Acknowledgments

References

[1] J. Lü, X. Yu, and G. Chen, "Chaos synchronization of general complex dynamical networks," *Physica A: Statistical Mechanics and its Applications*, vol. 334, no. 1-2, pp. 281–302, 2004.

[2] V. Perez-Munuzuri, V. Perez-Villar, and L. O. Chua, "Autowaves for image processing on a two-dimensional CNN array of excitable nonlinear circuits: flat and wrinkled labyrinths," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 40, no. 3, pp. 174–181, 1993.

[3] Y. Zhang and Z. He, "A secure communication scheme based on cellular neural network," in *1997 IEEE International Conference on Intelligent Processing Systems (Cat. No.97TH8335)*, pp. 521–524, Beijing, China, October 1997.

[4] V. Milanović and M. E. Zaghloul, "Synchronization of chaotic neural networks and applications to communications," *International Journal of Bifurcation and Chaos*, vol. 6, no. 12b, pp. 2571–2585, 1996.

[5] T. Kwok and K. A. Smith, "A unified framework for chaotic neural-network approaches to combinatorial optimization," *IEEE Transactions on Neural Networks*, vol. 10, no. 4, pp. 978–981, 1999.

[6] S. H. Strogatz and I. Stewart, "Coupled oscillators and biological synchronization," *Scientific American*, vol. 269, no. 6, pp. 102–109, 1993.

[7] Q. Song, "Design of controller on synchronization of chaotic neural networks with mixed time-varying delays," *Neurocomputing*, vol. 72, no. 13-15, pp. 3288–3295, 2009.

[8] Z. Cai, L. Huang, M. Zhu, and D. Wang, "Finite-time stabilization control of memristor-based neural networks," *Nonlinear Analysis: Hybrid Systems*, vol. 20, pp. 37–54, 2016.

[9] Z. Guo, J. Wang, and Z. Yan, "Global exponential synchronization of two memristor-based recurrent neural networks with time delays via static or dynamic coupling," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 2, pp. 235–249, 2015.

[10] Z. Guo, S. Yang, and J. Wang, "Global exponential synchronization of multiple memristive neural networks with time delay via nonlinear coupling," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 6, pp. 1300–1311, 2015.

[11] J. Cao, P. Li, and W. Wang, "Global synchronization in arrays of delayed neural networks with constant and delayed coupling," *Physics Letters A*, vol. 353, no. 4, pp. 318–325, 2006.

[12] J. Cao, G. Chen, and P. Li, "Global synchronization in an array of delayed neural networks with hybrid coupling," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 2, pp. 488–498, 2008.

[13] Z.-G. Wu, P. Shi, H. Su, and J. Chu, "Sampled-data exponential synchronization of complex dynamical networks with time-varying coupling delay," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 8, pp. 1177–1187, 2013.

[14] W. Lu, T. Chen, and G. Chen, "Synchronization analysis of linearly coupled systems described by differential equations with a coupling delay," *Physica D: Nonlinear Phenomena*, vol. 221, no. 2, pp. 118–134, 2006.

[15] W. He and J. Cao, "Exponential synchronization of hybrid coupled networks with delayed coupling," *IEEE Transactions on Neural Networks*, vol. 21, no. 4, pp. 571–583, 2010.

[16] P. Li, J. Cao, and Z. Wang, "Robust impulsive synchronization of coupled delayed neural networks with uncertainties," *Physica A: Statistical Mechanics and its Applications*, vol. 373, no. 1, pp. 261–272, 2007.

[17] S. Cai, J. Hao, Q. He, and Z. Liu, "Exponential synchronization of complex delayed dynamical networks via pinning periodically intermittent control," *Physics Letters A*, vol. 375, no. 19, pp. 1965–1971, 2011.

[18] X. F. Wang and G. Chen, "Pinning control of scale-free dynamical networks," *Physica A: Statistical Mechanics and its Applications*, vol. 310, no. 3-4, pp. 521–531, 2002.

[19] J. Zhou, J.-a. Lu, and J. Lü, "Pinning adaptive synchronization of a general complex dynamical network," *Automatica*, vol. 44, no. 4, pp. 996–1003, 2008.

[20] C. Yang and L. Huang, "Finite-time synchronization of coupled time-delayed neural networks with discontinuous activations," *Neurocomputing*, vol. 249, no. 8, pp. 64–71, 2017.

[21] J. Aubin and A. Cellina, "Differential inclusions," in *Differential inclusions*, pp. 8–13, Springer, 1984.

[22] J. P. la Salle, *The Stability of Dynamical Systems*, Society for Industrial and Applied Mathematics, 1976.

[23] A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides*, Springer Netherlands, Dordrecht, 1988.

[24] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Society for Industrial and Applied Mathematics, 1990.

[25] M. Forti and P. Nistri, "Global convergence of neural networks with discontinuous neuron activations," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 11, pp. 1421–1435, 2003.

[26] M. Forti, P. Nistri, and D. Papini, "Global exponential stability and global convergence in finite time of delayed neural networks with infinite gain," *IEEE Transactions on Neural Networks*, vol. 16, no. 6, pp. 1449–1463, 2005.

[27] Y. Yang and J. Cao, "Exponential synchronization of the complex dynamical networks with a coupling delay and impulsive effects," *Nonlinear Analysis: Real World Applications*, vol. 11, no. 3, pp. 1650–1659, 2010.

[28] W. Yu, J. Cao, and J. Lü, "Global synchronization of linearly hybrid coupled networks with time-varying delay," *SIAM Journal on Applied Dynamical Systems*, vol. 7, no. 1, pp. 108–133, 2008.

Research Article

More on Spectral Analysis of Signed Networks

Guihai Yu [1] and Hui Qu[1,2]

[1]*School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang, Guizhou, 550025, China*
[2]*Shandong Co-Innovation Center of Future Intelligent Computing, Yantai, Shandong, 264005, China*

Correspondence should be addressed to Guihai Yu; yuguihai@126.com

Spectral graph theory plays a key role in analyzing the structure of social (signed) networks. In this paper we continue to study some properties of (normalized) Laplacian matrix of signed networks. Sufficient and necessary conditions for the singularity of Laplacian matrix are given. We determine the correspondence between the balance of signed network and the singularity of its Laplacian matrix. An expression of the determinant of Laplacian matrix is present. The symmetry about 1 of eigenvalues of normalized Laplacian matrix is discussed. We determine that the integer 2 is an eigenvalue of normalized Laplacian matrix if and only if the signed network is balanced and bipartite. Finally an expression of the coefficient of normalized Laplacian characteristic polynomial is present.

1. Introduction

Social networks represent a large proportion of the complex socioeconomic organization in modern society which represent social entities including countries, corporations, or people. These entities interconnected through a wide range of social ties such as political treaties, commercial trade, friend, and collaboration. To display the ally/enemy, friend/foe, and trust/distrust relationships, the social system can be well represented by a signed network in which an edge of the network is assigned to be positive if two individuals are ally, friendship, trust, and negative if they are enemy, foe, and distrust. The origin of the study of signed networks can be tracked back to the work of Heider [1]. The use of signed networks was then proposed by Cartwright and Harary [2] to model the existence of balance/unbalance in the social networks.

As we know, graphs are very useful ways of presenting information about signed networks. However, when there are many actors and/or many kinds of relations, they can become so visually complicated that it is very difficult to see patterns. It is also possible to present information about signed networks in the form of matrices. Representing the information in this way also allows the application of mathematical and computer tools to summarize and find patterns.

Up to now, some matrices are employed by signed networks analysts in a number of different ways. This is the so-called *spectral graph theory*, which is a branch of mathematical science. Its idea is to exploit numerous relationship between the structure of a network (graph) and the spectrum of some matrix (or collection of matrices) associated with the network (graph). There are many different matrices that are employed, including adjacency matrix, Laplacian matrix, and normalized Laplacian matrix. The goal of this paper is to investigate some properties of Laplacian matrix and normalized Laplacian matrix of signed networks and exploit some relation between these matrices and signed networks.

Let G be an undirected network of order n with vertex set $V(G) = \{v_1, v_2, \ldots, v_n\}$ and edge set $E(G)$. The adjacency matrix $A(G) = (a_{ij})_{n \times n}$ of G is defined as follows: $a_{ij} = 1$ if v_i and v_j are adjacent and $a_{ij} = 0$ otherwise. A signed network $\Gamma = (G, \sigma)$ consists of a network $G = (V, E)$, referred to as its underlying network, and a sign function $\sigma : E \longrightarrow \{+, -\}$. The *adjacency matrix* of Γ is $A(\Gamma) = (a_{ij}^\sigma)$ with $a_{ij}^\sigma = \sigma(v_i v_j) a_{ij}$, where a_{ij} is an element in the adjacency matrix of the underlying network G and $v_i v_j$ is an edge of G. If all edges are signed positive, the adjacency matrix $A(G, \sigma)$ is exactly the ordinary adjacency matrix $A(G)$. Let $D(\Gamma) = \text{diag}(d_1, d_2, \ldots, d_n)$ be a diagonal matrix where

d_i is the degree of vertex v_i in its underlying network. The *Laplacian matrix* of Γ, denoted by $L(\Gamma)$, is defined as $D(\Gamma) - A(\Gamma)$. The matrix $D^{-1/2}L(\Gamma)D^{-1/2}$ is said to be *normalized Laplacian matrix* of Γ, denoted by $\mathscr{L}(\Gamma)$.

A signed i_1-i_k-walk W in a signed network Γ is a sequence of vertices and edges $W : v_{i_1}e_{12}v_{i_2}e_{23}v_{i_3}\cdots v_{i_{k-1}}e_{(k-1)k}v_{i_k}$ such that $e_{s(s+1)} = v_{i_s}v_{i_{s+1}} \in E(\Gamma)$ $(s = 1, 2, \ldots, k-1)$. An i_1-i_k-walk W is called *even* (odd) if k is even (odd). The sign of a signed walk $W = v_1e_{12}v_2e_{23}\cdots v_l$ is $\text{sgn}(W) = a^\sigma_{12}a^\sigma_{23}\cdots a^\sigma_{(l-1)l}$ and $e_{i(i+1)} = v_iv_{i+1}$ $(i = 1, 2, \ldots, l-1)$. A signed walk W is *balanced* (*unbalanced*) if $\text{sgn}(W) = 1$ ($\text{sgn}(W) = -1$). A signed cycle is called *balanced* (unbalanced) if its sign is +1 (−1). A signed networks is called *balanced (resp. unbalanced)* if each its signed cycle is balanced (resp. unbalanced).

Suppose that $\Gamma = (G, \sigma)$ is a signed network. A signed function $\theta : V \longrightarrow \{+1, -1\}$ is a switching function if Γ is transformed to a new signed network $\Gamma^\theta = (G, \sigma^\theta)$ by θ such that the underlying graph remains the same and the sign function is defined by $\sigma^\theta(e) = \theta(v_i)\sigma(e)\theta(v_j)$ for an edge $e = v_iv_j$. Let $\Gamma_1 = (G, \sigma_1)$ and $\Gamma_2 = (G, \sigma_2)$ be two signed networks with the same underlying graph. We call Γ_1 and Γ_2 *switching equivalent* and write $\Gamma_1 \sim \Gamma_2$, if there exists a switching function θ such that $\Gamma_2 = \Gamma_1^\theta$. Switching preserves some signed-graphic invariants such as the sign of cycles and spectrum of combinatorial matrices (adjacency matrix, normalized Laplacian matrix).

This paper is organized as follows. In Section 2, we study some properties of Laplacian matrix of signed networks. Sufficient and necessary conditions for the singularity of Laplacian matrix are given. The correspondence between the balance of signed network and the singularity of its Laplacian matrix is determined. An expression of the determinant of Laplacian matrix is present. In Section 3, the symmetry about 1 of eigenvalues of normalized Laplacian matrix is discussed. Sufficient and necessary condition for that the integer 2 is an eigenvalue of normalized Laplacian matrix is given. An expression of all coefficients of normalized Laplacian characteristic polynomial is present.

2. Laplacian Matrix and Signed Network

Hou et al. [3] introduced the incidence matrix of a signed network as follows. Let $S(\Gamma) = (s_{ij})$ be an $n \times m$ matrix indexed by the vertex and the edge of signed network Γ and

$$s_{ij} = \begin{cases} +1 & \text{if } v_i \text{ is the head of } e_j \\ +1 & \text{if } v_i \text{ is the tail of } e_j \text{ and } \sigma\left(e_j\right) = + \\ -1 & \text{if } v_i \text{ is the tail of } e_j \text{ and } \sigma\left(e_j\right) = - \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The following is immediate by the direct calculation.

Theorem 1 (see [3]). *Let Γ be a signed network. Then $L(\Gamma) = S(\Gamma)S^T(\Gamma)$ and $L(\Gamma)$ is a positive semidefinite matrix.*

Theorem 2. *Let Γ be a connected signed network on vertices $v_1, v_2, \ldots, v_n$. Then $L(\Gamma)$ is singular if and only if any 1-i-walk*

has the same sign. In this case, 0 is a simple eigenvalue with an eigenvector $\alpha = (1, \text{sgn}(W_2), \text{sgn}(W_3), \ldots, \text{sgn}(W_n))^T$, where W_i is a 1-i-walk in Γ.

Proof. Let $x^T = (x_1, x_2, \ldots, x_n) \in C^n$. Note that for any nonzero vector x, $L(\Gamma) = 0$ if and only if $S^T(\Gamma)x = 0$. By (14), $S^T(\Gamma)x = 0$ if and only if $x_i = a^\sigma_{ij}x_j$ for any edge $e = v_iv_j$. Let $W_i = u_1u_2\cdots u_i$ be any 1-i-walk and $u_1 = v_1, u_i = v_i$. Suppose that $L(\Gamma) = 0$. So we have

$$\begin{aligned} x_1 &= a^\sigma_{12}x_2 = a^\sigma_{12}a^\sigma_{23}x_3 = \cdots = a^\sigma_{12}a^\sigma_{23}\cdots a^\sigma_{(i-1)i}x_i \\ &= \text{sgn}\left(W_i\right)x_i, \end{aligned} \quad (2)$$

which implies that each 1-i-walk has the same sign.

Note that $\text{sgn}^{-1}(W_i) = \text{sgn}(W_i)$. Hence

$$\begin{aligned} x^T &= (x_1, x_2, \ldots, x_n) \\ &= (x_1, \text{sgn}\left(W_2\right)x_1, \text{sgn}\left(W_3\right)x_1, \ldots, \text{sgn}\left(W_n\right)x_1) \quad (3) \\ &= x_1\alpha^T. \end{aligned}$$

This implies that 0 is a simple eigenvalue of $L(\Gamma)$ with an eigenvector α.

Suppose that any 1-i-walk has the sign. Then for any edge $e_{ij} \in E(\Gamma)$, $\text{sgn}(W_j) = \text{sgn}(W_i) \cdot a^\sigma_{ij}$. Let $x^T = (x_1, x_2, \ldots, x_n)$ be a column vector such that $x_i = \text{sgn}(W_i)x_1$ $(i = 2, 3, \ldots, n)$. Then $x_j = \text{sgn}(W_j)x_1 = \text{sgn}(W_i) \cdot a^\sigma_{ij}x_1 = x_ia^\sigma_{ij}$, i.e., $x_i = a^\sigma_{ij}x_j$. So we have

$$\begin{aligned} x^T L\left(\Gamma\right)x &= \sum_{e_{ij} \in E(\Gamma)} \left|x_i - a^\sigma_{ij}x_j\right|^2 = \sum_{e_{ij} \in E(\Gamma)} \left|a^\sigma_{ij}x_j - a^\sigma_{ij}x_j\right|^2 \\ &= 0. \end{aligned} \quad (4)$$

This implies that $L(\Gamma)$ is singular. $\square$

Note that for the underlying network it is known that the multiplicity of the eigenvalue 0 of Laplacian matrix is equal to the number of components. For signed network, the following holds from the proof of Theorem 2.

Theorem 3. *The multiplicity of the eigenvalue 0 of Laplacian matrix of a signed network is the number of components whose Laplacian matrix is singular.*

Theorem 4 (see [4]). *A signed network is balanced if and only if for each pair of distinct vertices v_1, v_2 all paths joining v_1 and v_2 have the same sign.*

From Theorems 2 and 4, we have the following.

Theorem 5. *A signed network Γ is balanced if and only if $L(\Gamma)$ is singular.*

The following is immediate from Theorem 5.

Theorem 6. *The Laplacian matrix of a signed network is singular if and only if the Laplacian matrix of any its cycles is singular. In particular, the Laplacian matrix of any acyclic graph is singular.*

In [5], authors determined the determinant of the Laplacian matrix of mixed graphs. Here by the similar method we shall extend it to the case for signed graphs.

Theorem 7. $\det L(C) = 2[1 - \mathrm{sgn}(C)]$ *for any signed cycle* C.

Proof. Let C be a signed cycle with vertex set $V(C) = \{v_1, v_2, \ldots, v_n\}$ and edge set $E(C) = \{e_1, e_2, \ldots, e_n\}$ such that $e_i = v_i v_{i+1}$ $(1 \le i \le n - 1)$ and $e_n = v_n v_1$. For the incidence matrix $S(C)$, we expand its the first row

$$\det S(C) = \prod_{i=1}^{n} s_{ie_i} + (-1)^{n+1} s_{1e_n} \prod_{i=2}^{n} s_{ie_{i-1}}. \tag{5}$$

By directly calculation and the fact that $s_{ie_i} s_{je_j} = -\sigma(e)a_{ij} = -a_{ij}^{\sigma}$ for any edge $e_i = v_i v_{i+1}$. It follows that

$$\det L(C) = \det S(C) \cdot \det S^T(C) = 2 - 2\,\mathrm{sgn}(C). \tag{6}$$

So the result holds. □

Theorem 8. *Let* Γ *be a signed unicyclic network with a cycle* C. *Then*

$$\det L(\Gamma) = \det L(C) = 2[1 - \mathrm{sgn}(C)]. \tag{7}$$

Proof. By Theorem 7, the results hold if Γ is a signed cycle. Assume that Γ has a pendant vertex, say u. Let v be the unique neighbor of u in Γ. Let e be the edge joining u and v. After permutations, the first row and the first column of $S(\Gamma)$ correspond to the vertex u and the edge e, respectively. Note that $S(\Gamma)$ is a square matrix since Γ is unicyclic. We get the determinant of $S(\Gamma)$ by expanding along the first row as follows:

$$\det S(\Gamma) = s_{ue} \cdot \det S(\Gamma'), \tag{8}$$

where Γ' is a signed subgraph obtained from Γ by deleting the vertex u. Hence we have

$$\det L(\Gamma) = \det S(\Gamma) \cdot \det S^T(\Gamma)$$
$$= \det S(\Gamma') \cdot \det S^T(\Gamma') = \det L(\Gamma'). \tag{9}$$

Repeating the above finite steps, we have $\det L(\Gamma) = \det L(C)$. □

Let Γ be a connected signed network. We call a subnetwork H as *an essential spanning subnetwork* of Γ if either Γ is balanced and H is a spanning tree of Γ, or else Γ is not balanced, $V(\Gamma) = V(H)$ and every component of H is a unicyclic signed network in which the unique cycle is negative. By $\mathscr{E}(\Gamma)$ we denote the set of all essential spanning subnetworks of Γ.

Theorem 9. *Let* Γ *be a connected signed network. Then*

$$\det L(\Gamma) = \sum_l 4^l b_l, \tag{10}$$

where b_l *is the number of essential spanning subgraphs which contain* l *unbalanced cycles and* $b_0 = 0$.

Proof. It is evident that the result holds if Γ is a tree. Assume that Γ contains some cycles. By Cauchy-Binet Theorem [6] and $L(\Gamma) = S(\Gamma) \cdot S^T(\Gamma)$, we have

$$\det L(\Gamma) = \sum_{E' \subseteq E(\Gamma); |E'|=|V(\Gamma)|} \det S[V(\Gamma), E']$$
$$\cdot \det S^T[V(\Gamma), E'], \tag{11}$$

where $S[V(\Gamma), E']$ is a square submatrix of $S(\Gamma)$.

Note that $S[V(\Gamma), E']$ is the vertex-edge incidence matrix of a spanning subgraph of Γ, say $H_{E'}$, with the edge set $|E'| = |V(\Gamma)|$. Moreover, $\det L(H_{E'}) = S[V(\Gamma), E'] \cdot S^T[V(\Gamma), E']$. Note that every component of $H_{E'}$ is unicyclic and $H_{E'} \in \mathscr{E}(\Gamma)$. By Theorem 8, we have

$$\det L(\Gamma) = \sum_{E' \subseteq E(\Gamma); |E'|=|V(\Gamma)|} \det S[V(\Gamma), E']$$
$$\cdot \det S^T[V(\Gamma), E']$$
$$= \sum_{E' \subseteq E(\Gamma); |E'|=|V(\Gamma)|} \det L(H_{E'})$$
$$= \sum_{H \in \mathscr{E}(\Gamma)} \det L(H) \tag{12}$$
$$= \sum_{H \in \mathscr{E}(\Gamma)} \prod_{i=1}^{b_l} 2(1 - \mathrm{sgn}(C_i(H)))$$
$$\text{where } \mathrm{sgn}(C_i(H)) = -1 = \sum_{l=0} 4^l b_l.$$

So the result holds. □

The following is immediate from Theorem 9, which is coincident with the definition of balance of signed network.

Theorem 10. *Let* Γ *be a signed network. Then* Γ *is balanced if and only if each cycle of* Γ *is balanced cycle.*

3. Normalized Laplacian Matrix and Signed Network

For a signed network Γ, the normalized Laplacian matrix $\mathscr{L}(\Gamma)$ is symmetric and positive semidefinite [7], so its eigenvalues are real and nonnegative, denoted by $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Firstly we recall some properties of normalized Laplacian matrix.

Lemma 11 (see [7]). *Let* Γ *be a signed network on n vertices with normalized Laplacian eigenvalues* $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. *Then* $\lambda_n \le 2$.

Lemma 12 (see [3, 7]). *Let* Γ_1 *and* Γ_2 *be two signed networks with the same underlying network. Then* $\Gamma_1 \sim \Gamma_2$ *if and only if* $\mathscr{L}(\Gamma_1)$ *and* $\mathscr{L}(\Gamma_2)$ *are signature similar.*

In [8], the symmetry about 1 of eigenvalues for bipartite signed network was present as follows. Here we present a stronger result.

Theorem 13 (see [8]). *Let $\Gamma = (G, \sigma)$ be a bipartite signed network. If λ is an eigenvalue of $\mathscr{L}(\Gamma)$, then $2 - \lambda$ is also an eigenvalue of $\mathscr{L}(\Gamma)$.*

Theorem 14. *Let Γ be a connected signed network. Then Γ is bipartite if and only if all eigenvalues of $\mathscr{L}(\Gamma)$ are symmetric about 1 (including multiplicities); i.e., for each eigenvalue λ_i, $2 - \lambda_i$ is also an eigenvalue of $\mathscr{L}(\Gamma)$.*

Proof. It suffices to verify that $I - \mathscr{L}(\Gamma)$ and $-(I - \mathscr{L}(\Gamma))$ have the same spectrum. Note that $I - \mathscr{L}(\Gamma) = D^{-1/2}A(\Gamma)D^{-1/2}$. Γ is bipartite if and only if $D^{-1/2}A(\Gamma)D^{-1/2}$ can be expressed as $\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$. It is evident that

$$\begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & -B \\ -B^T & 0 \end{pmatrix} \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}. \quad (13)$$

This yields to the result. □

From Lemma 11, the integer 2 is the upper bound of normalized Laplacian eigenvalues. In this sequel, we give a sufficient and necessary condition for that the integer 2 is an eigenvalue of normalized Laplacian matrix.

Theorem 15. *Let Γ be a connected signed network. Then 2 is an eigenvalue of $\mathscr{L}(\Gamma)$ if and only if Γ is a balanced bipartite signed network.*

Proof. By Courant-Fischer theorem, we have

$$\lambda_n = \sup_{f \neq 0} \frac{\sum_{u \sim v} \left(f(u) - \sigma(uv) f(v) \right)^2}{\sum_v f^2(v) d(v)}. \quad (14)$$

Assume that 2 is an eigenvalue of $\mathscr{L}(\Gamma)$ with nonzero eigenvector $y^T = (y_1, y_2, \ldots, y_n)$. By Lemma 11 and (14), $y_i = -\sigma(v_i v_j) y_j$ for any edge e incident to v_i and v_j. So $V(\Gamma)$ can be partitioned into two parts such that no edge existing between any two vertices in every part. This means that Γ is bipartite. For any even cycle $C_{2k} = v_1 v_2 \cdots v_{2k} v_1$, we have

$$y_1 = -\sigma(v_1 v_2) y_2 = \sigma(v_1 v_2) \sigma(v_2 v_3) y_3$$

$$= -\sigma(v_1 v_2) \sigma(v_2 v_3) \sigma(v_3 v_4) y_4 = \cdots \quad (15)$$

$$= -\sigma(v_1 v_2) \sigma(v_2 v_3) \cdots \sigma(v_{2k-1} v_{2k}) y_{2k}.$$

Moreover, $y_1 = -\sigma(v_1 v_{2k}) y_{2k}$. So $\sigma(v_1 v_2)\sigma(v_2 v_3) \cdots \sigma(v_{2k-1} v_{2k}) = \sigma(v_1 v_{2k})$ and C_{2k} is balanced. This implies that Γ is balanced.

If Γ is balanced bipartite, then 0 is an eigenvalue of $\mathscr{L}(\Gamma)$. By Theorem 14 and Lemma 12, 2 is an eigenvalue of $\mathscr{L}(\Gamma)$. □

As we know, the coefficients of characteristic polynomial of adjacency (Laplacian) matrix are related to the graph structure. In [9], expressions of coefficients of (Laplacian) characteristic polynomial was present. We would present the expression of the coefficients of normalized Laplacian characteristic polynomial. Firstly, we recall the Sachs formula for the coefficients of adjacency characteristic polynomial of signed networks. Here some definitions are needed. An *elementary figure* is the graph K_2 or the cycle. A *basic figure* is the disjoint union of elementary figures.

Lemma 16 (see [9]). *Let $\Gamma = (G, \sigma)$ and $\phi(\Gamma, x) = x^n + a_1 x^{n-1} + \cdots + a_n$ be a signed network and its adjacency characteristic polynomial, respectively. Then*

$$a_i = \sum_{B \in \mathscr{B}_i} (-1)^{p(B)} 2^{|c(B)|} \sigma(B), \quad (16)$$

where $\mathscr{B}_i$ is the set of basic figures on i vertices in G, $p(B)$ is the number of components of B, and $c(B)$ is the set of cycles in B and $\sigma(B) = \prod_{C \in c(B)} \text{sgn}(C)$.

Let $\psi(\Gamma, x)$ be the normalized Laplacian characteristic polynomial of Γ. By the definition of normalized Laplacian matrix, we have

$$\psi(\Gamma, x) = \det(xI - \mathscr{L}(\Gamma))$$

$$= \det\left(xI - D^{-1/2}L(\Gamma)D^{-1/2}\right)$$

$$= \det\left(xI - I + D^{-1/2}A(\Gamma)D^{-1/2}\right) \quad (17)$$

$$= \det\left((x-1)I + D^{-1/2}A(\Gamma)D^{-1/2}\right)$$

$$= (x-1)^n + c_1(x-1)^{n-1} + \cdots + c_{n-1}(x-1)$$

$$+ c_n.$$

Theorem 17. *Let Γ be a signed network on n vertices and $\psi(\Gamma, x) = (x-1)^n + c_1(x-1)^{n-1} + \cdots + c_{n-1}(x-1) + c_n$ be its normalized Laplacian characteristic polynomial. Then*

$$c_k = \sum_{B \in \mathscr{B}_k} (-1)^{p(B)} 2^{|c(B)|} \sigma(B) \frac{1}{D_k}, \quad (18)$$

where $\mathscr{B}_k$ is the set of basic figures on k vertices in G, $p(B)$ is the number of components of B, $c(B)$ is the set of cycles in B, $\sigma(B) = \prod_{C \in c(B)} \text{sgn}(C)$, $D_k = \prod_{v_i \in V(B)} d_i$, and d_i is the degree of v_i in Γ.

Proof. Note that

$$\psi(\Gamma, x) = \det\left((x-1)I + D^{-1/2}A(\Gamma)D^{-1/2}\right)$$

$$= (-1)^n \det\left((1-x)I - D^{-1/2}A(\Gamma)D^{-1/2}\right). \quad (19)$$

Set

$$\det\left((1-x)I - D^{-1/2}A(\Gamma)D^{-1/2}\right)$$

$$= (x-1)^n + c_1'(x-1)^{n-1} + \cdots + c_{n-1}'(x-1) + c_n'. \quad (20)$$

So $c_k = (-1)^k c_k'$. Moreover, $(-1)^k c_k'$ equals to the sum of all $k \times k$ minors of $D^{-1/2}A(\Gamma)D^{-1/2}$. Then c_k is the sum of all $k \times k$ minors of $D^{-1/2}A(\Gamma)D^{-1/2}$. It is evident that each such $k \times k$ minor of $D^{-1/2}A(\Gamma)D^{-1/2}$ is the product of the corresponding $k \times k$ minors of $D^{-1/2}$, $A(\Gamma)$, and $D^{-1/2}$, respectively. Furthermore, any $k \times k$ minor of $A(\Gamma)$ is the determinant of adjacency matrix of an induced subgraph of Γ with k vertices. So this result holds from Lemma 16. □

4. Conclusion

Recently, there are some results on the spectral theory of signed graphs [10–18]. In this paper we investigate some properties of (normalized) Laplacian matrix of signed network and present a correspondence between the balance of signed networks and the singularity of Laplacian matrix. Moreover, we give the expressions of determinant of Laplacian matrix and coefficients of normalized Laplacian characteristic polynomial, respectively. Actually there are some other aspects of spectrum of signed graphs, which can be investigated. It will be left to our future study. In addition, there are many spectrum-based invariants, which are widely investigated, such as graph energy (e.g., graph theory [19, 20], incidence energy [21], and matching energy [22, 23]), HOMO-LUMO index [24, 25], and inertia [26–29]. In the future, we would like to study some properties of these spectrum-based indices of signed networks.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] F. Heider, "Attitude and cognitive organization," *The Journal of Psychology*, vol. 21, pp. 107–122, 1946.

[2] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory," *Psychological Review*, vol. 63, no. 5, pp. 277–293, 1956.

[3] Y. Hou, J. Li, and Y. Pan, "On the Laplacian eigenvalues of signed graphs," *Linear and Multilinear Algebra*, vol. 51, no. 1, pp. 21–30, 2003.

[4] F. Harary, "On the notion of balance of a signed graph," *Michigan Mathematical Journal*, vol. 2, no. 2, pp. 143–146, 1953.

[5] G. Yu, X. Liu, and H. Qu, "Singularity of Hermitian (quasi-)Laplacian matrix of mixed graphs," *Applied Mathematics and Computation*, vol. 293, pp. 287–292, 2017.

[6] A. H. Roger and R. J. Charles, *Matrix Analysis*, Cambridge University Press, New York, NY, USA, 2nd edition, 2013.

[7] H.-H. Li and J.-S. Li, "Note on the normalized Laplacian eigenvalues of signed graphs," *The Australasian Journal of Combinatorics*, vol. 44, pp. 153–162, 2009.

[8] F. M. Atay and H. Tunçel, "On the spectrum of the normalized Laplacian for signed graphs: interlacing, contraction, and replication," *Linear Algebra and Its Applications*, vol. 442, pp. 165–177, 2014.

[9] F. Belardo and S. K. Simic', "On the Laplacian coefficients of signed graphs," *Linear Algebra and Its Applications*, vol. 475, pp. 94–113, 2015.

[10] Y.-Z. Fan, Y. Wang, and Y. Wang, "A note on the nullity of unicyclic signed graphs," *Linear Algebra and Its Applications*, vol. 438, no. 3, pp. 1193–1200, 2013.

[11] Y.-Z. Fan, W.-X. Du, and C.-L. Dong, "The nullity of bicyclic signed graphs," *Linear and Multilinear Algebra*, vol. 62, no. 2, pp. 242–251, 2014.

[12] I. Gutman, S.-L. Lee, J.-H. Sheu, and C. Li, "Predicting the, nodal properties of molecular orbitals by means of signed graphs," *Bulletin of the Institute of Chemistry, Academia Sinica*, vol. 42, pp. 25–32, 1995.

[13] Y. P. Hou, "Bounds for the least Laplacian eigenvalue of a signed graph," *Acta Mathematica Sinica*, vol. 21, no. 4, pp. 955–960, 2005.

[14] S. Lee and C. Li, "Chemical signed graph theory," *International Journal of Quantum Chemistry*, vol. 49, no. 5, pp. 639–648, 1994.

[15] S.-L. Lee, R. R. Lucchese, and S. Y. Chu, "Topological analysis of eigenvectors of the adjacency matrices in graph theory: the concept of internal connectivity," *Chemical Physics Letters*, vol. 137, no. 3, pp. 279–284, 1987.

[16] P. K. Sahu and S.-L. Lee, "Net-sign identity information index: A novel approach towards numerical characterization of chemical signed graph theory," *Chemical Physics Letters*, vol. 454, no. 1-3, pp. 133–138, 2008.

[17] Y. Liu and L. You, "Further results on the nullity of signed graphs," *Journal of Applied Mathematics*, vol. 2014, Article ID 483735, 8 pages, 2014.

[18] S. Wasserman and K. Faust, *Social Networks Analysis: Methods And Applications*, Cambridge University Press, Cambridge, 1994.

[19] R. Lang, T. Li, D. Mo, and Y. Shi, "A novel method for analyzing inverse problem of topological indices of graphs using competitive agglomeration," *Applied Mathematics and Computation*, vol. 291, pp. 115–121, 2016.

[20] X. Li, Y. Shi, and I. Gutman, *Graph Theory*, Springer, New York, NY, USA, 2012.

[21] S. B. Bozkurt and D. Bozkurt, "On incidence energy," *MATCH - Communications in Mathematical and in Computer Chemistry*, vol. 72, no. 1, pp. 215–225, 2014.

[22] L. Chen, J. Liu, and Y. Shi, "Bounds on the matching energy of unicyclic odd-cycle graphs," *MATCH - Communications in Mathematical and in Computer Chemistry*, vol. 75, no. 2, pp. 315–330, 2016.

[23] L. Chen, J. Liu, and Y. Shi, "Matching energy of unicyclic and bicyclic graphs with a given diameter," *Complexity*, vol. 21, no. 2, pp. 224–238, 2015.

[24] X. Li, Y. Li, Y. Shi, and I. Gutman, "Note on the HOMO-LUMO index of graphs," *MATCH - Communications in Mathematical and in Computer Chemistry*, vol. 70, no. 1, pp. 85–96, 2013.

[25] B. Mohar, "Median eigenvalues of bipartite planar graphs," *MATCH - Communications in Mathematical and in Computer Chemistry*, vol. 70, no. 1, pp. 79–84, 2013.

[26] G. Yu, H. Qu, and J. Tu, "Inertia of complex unit gain graphs," *Applied Mathematics and Computation*, vol. 265, pp. 619–629, 2015.

[27] G. Yu, L. Feng, and Q. Wang, "Bicyclic graphs with small positive index of inertia," *Linear Algebra and its Applications*, vol. 438, no. 5, pp. 2036–2045, 2013.

[28] G. Yu, L. Feng, Q. Wang, and A. Ilic, "The minimal positive index of inertia of signed unicyclic graphs," *Ars Combinatoria*, vol. 117, pp. 245–255, 2014.

[29] G. Yu, X.-D. Zhang, and L. Feng, "The inertia of weighted unicyclic graphs," *Linear Algebra and its Applications*, vol. 448, pp. 130–152, 2014.

Research Article

Competition-Based Benchmarking of Influence Ranking Methods in Social Networks

Alexandru Topîrceanu [ID]

Department of Computer and Information Technology, Politehnica University Timisoara, Timisoara, Romania

Correspondence should be addressed to Alexandru Topîrceanu; alext@cs.upt.ro

The development of new methods to identify influential spreaders in complex networks has been a significant challenge in network science over the last decade. Practical significance spans from graph theory to interdisciplinary fields like biology, sociology, economics, and marketing. Despite rich literature in this direction, we find small notable effort to consistently compare and rank existing centralities considering both the topology and the opinion diffusion model, as well as considering the context of *simultaneous* spreading. To this end, our study introduces a new benchmarking framework targeting the scenario of *competitive opinion diffusion*; our method differs from classic SIR epidemic diffusion, by employing competition-based spreading supported by the realistic tolerance-based diffusion model. We review a wide range of state-of-the-art node ranking methods and apply our novel method on large synthetic and real-world datasets. Simulations show that our methodology offers much higher quantitative differentiation between ranking methods on the same dataset and notably high granularity for a ranking method over different datasets. We are able to pinpoint—with consistency—which influence the ranking method performs better against the other one, on a given complex network topology. We consider that our framework can offer a forward leap when analysing diffusion characterized by real-time competition between agents. These results can greatly benefit the tackling of social unrest, rumour spreading, political manipulation, and other vital and challenging applications in social network analysis.

1. Introduction

Estimating node influence can lead to an improved understanding of the natural interaction patterns within real-world populations, biological entities, or technological structures. The applicability of metrics for quantifying the influence potential of nodes has wide-ranging interdisciplinary applications including disease modelling [1–7], information transmission [8–11], behavioural intelligence [3, 12–15], business management [16, 17], finances [18, 19], and pharmacology and drug repurposing [20, 21]. Being able to correctly determine and rank influential nodes in empirical networks can have direct applicability in problems like impeding epidemic outbreaks [22], accelerating innovation diffusion, evaluating marketing and financial trends [18], discovering new drug targets in pathway networks [20], and

predicting essential proteins in protein interaction networks and gene regulatory networks [23]. Regardless of the context, the most common way to capture information on intricate real-world interactions is a complex network [24–27]. Specifically, social network analysis (SNA), as a subdomain of network science, models social structures characterized by emergent interaction.

There is considerable effort devoted to assessing the importance of nodes in many types of complex networks over the last decade. Novel approaches, combined with classic graph centrality measures, have led to the emergence of the three main categories of influence ranking methods. The first category of scientists argues that the location of a node is more important than its immediate ego network and thus proposed k-core decomposition [28, 29], along with improved variants, such as [30–33]. The second category of
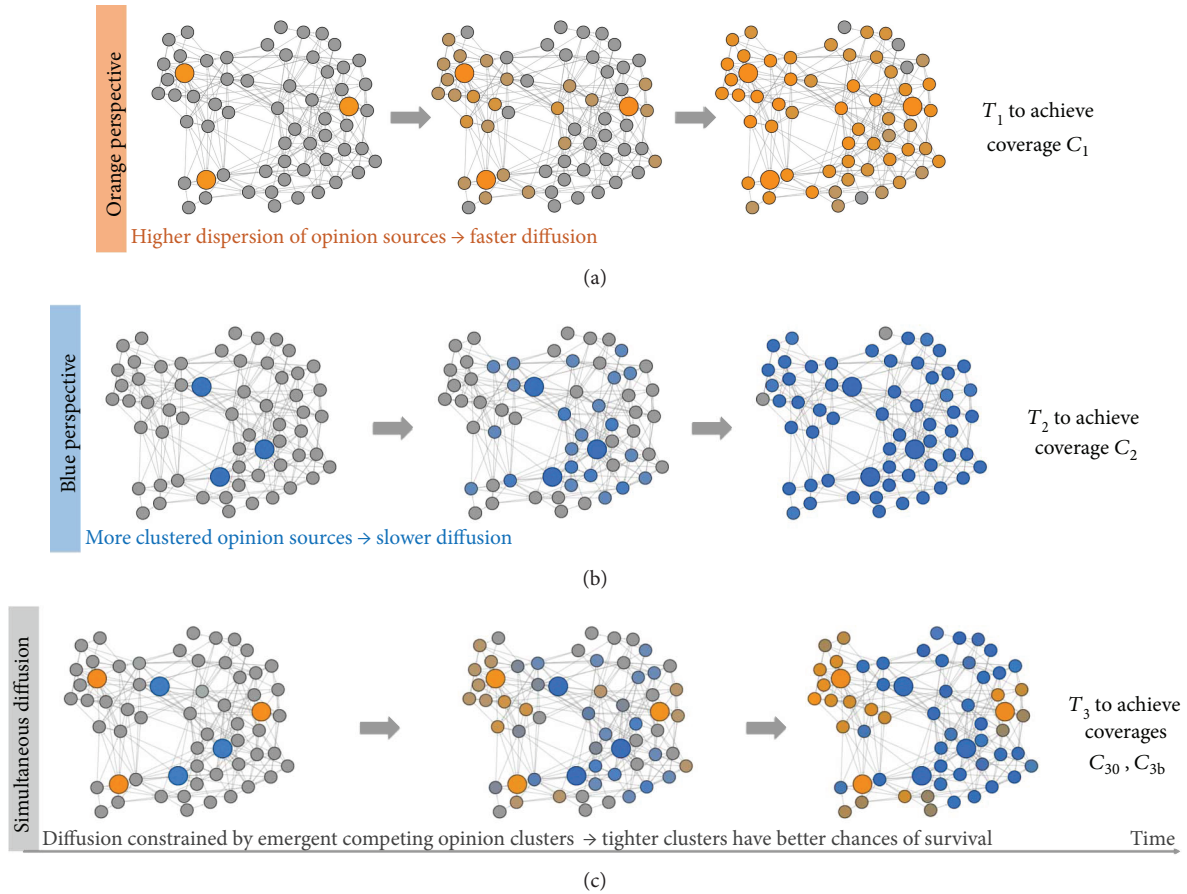
FIGURE 1: Example of the incurred limitations when benchmarking a diffusion process only from a single opinion's point of view, when the real-world context implies simultaneous diffusion and competition between multiple opinions. It is suggested that the orange diffusion time T_1 is shorter (better) than the blue T_2 time, due to the higher, more uniform dispersion of orange opinion sources. However, in reality (c), none of the two opinions may fully cover the network in optimal times T_1 or T_2, nor will they achieve such high coverages as C_1 or C_2, that is, $T_1 \approx T_2 < T_3$ and $(C_1 \approx C_2) > C_{3o}, C_{3b}$.

scientists quantifies the influence of a node based solely on its local surroundings [34–36]. Finally, the third category of scientists evaluates node influences according to various states of equilibrium for dynamical processes, such as random walks [37, 38] or step-wise refinements [39].

Each ranking method, regardless of its nature and category, is validated through a state-of-the-art benchmarking methodology, which—in almost all cases of network science—involves the usage of the SIR epidemic model [40–42]. This process may be suitable for validating metrics in an individual context in order to produce a verdict whether the ranking method is good enough, but often not more. For SNA, however, collective interplay is inherent [43] and the aforementioned real-world application contexts imply competition between multiple opinions, so a one-sided perspective will often not be reliable. The recent study shows that the traditional SIR model provides a poor description of the data for modelling disease dynamics, as it lacks infectious recovery dynamics, which is a better description of social network dynamics [44]. Consequently, we consider that the SIR model would be inadequate to apply in our benchmarking context, as it fails to model competition and opinion

fluctuations. As such, we propose a more robust benchmarking principle that implies *simultaneous competition* between two or more information (opinion, rumour) sources, that is, in the same network and at the same time. To this end, we make use of the existing tolerance-based diffusion model [45], which represents, to the best of our knowledge, a novel benchmarking methodology in SNA.

To better underline the limitations incurred by using a SIR simulation versus our proposed competition-based benchmark, we illustrate a comparative example in Figure 1. In (a) and (b), we apply two distinct ranking methods (orange and blue), one at a time, and show that the diffusion process is unrestrained, also we suggest that orange manages to cover the network in time T_1, faster than blue with T_2, due to the higher dispersion of three initial orange opinion sources. In the SIR context, the two simulations may lead to the conclusion that the orange ranking method is better than the blue one. In reality, we consider the scenario in (c) as the more probable one. Opinions will diffuse simultaneously and face constraints due to competition over each node (i.e., orange and blue exclude one another). In this case, we intuitively suggest that blue might win in terms of

network coverage, as it has a tighter initial cluster forming around its three opinion sources. Consequently, the main observations are the following:

(i) None of the two opinions will achieve coverages as high as in the one-sided scenarios, that is, C_{3o} and $C_{3b} < (C_1 \approx C_2)$.

(ii) Simulation time T_3 may be longer than $T_1 \approx T_2$, due to the need for attaining a state of balance in the emergent network.

(iii) The final ratio of opinion C_{3o}/C_{3b} is impossible to determine by one-sided simulations and is only determinable by the emergence of the two competing opinions (e.g., initial spreader position, connectivity of the spreaders, and community structure).

In light of these remarks, we propose a novel benchmarking framework which offers more reliable insights into comparing ranking methods aimed at real-world applications of social networks. The paper starts by presenting the benchmarking methodology in detail, followed by simulation results. We highlight the overlapping of several popular ranking methods, in terms of selecting the same initial seeds, then proceed to compare the ranking methods using SIR as a reference and then in pairs (one versus one) using our proposed methodology. Finally, we discuss the results, the difference in what our testing methodology can offer, and what are the implications of considering competing opinion. The Methods section details the used validation datasets and a brief review of current state-of-the-art ranking measures used in complex networks.

2. A Novel Competition-Based Influence Ranking Benchmark

State-of-the-art benchmarking methodologies for spreading processes on complex networks often rely on the SIR (SIS) model [40–42]. With this approach, an initial subset of nodes is infected according to a centrality measure, then the simulation measures how fast surrounding susceptible nodes become recovered (i.e., including dead). Indeed, if we take the example of an epidemic, it spreads independently from other epidemics and has its own temporal evolution. On the other hand, if we consider opinion between social agents, it is often exclusive (in regard to other contradicting opinions) and is also dependent on the timing with the spread of other ideas.

We argue that a SIR model cannot accurately model *fluctuations* and *direct* competition between social agents. Also, as long as the infected nodes survive, they will eventually tamper with the whole network. Finally, the SIR model is sensitive to initial parameters, like infectious probability λ and recovery duration δ, needing step-wise refinements to obtain desired results, which may vary easily in other experimental settings. Alternatively, we find several variants of the SIR model designed for competitive diffusion processes, such as the SI_1I_2S [5], $SI_{1|2}S$ [6], and SI_1SI_2S [7] models, but they are targeting competitive epidemic diffusion.

As a novel, more robust, and more realistic alternative, we propose the usage of the tolerance-based model [45] which implies competition between two or more opinion sources in the same network, at the same time. To the best of our knowledge, this kind of benchmarking methodology is novel to literature. Other graph-based predictive diffusion models [46] include the classic linear threshold LT [47], independent cascade IC [48], voter model [49], Axelrod model [50], and Sznajd model [51]. These models use either fixed thresholds or thresholds evolving according to simple probabilistic processes that are not driven by the internal state of the social agents [46]. However, the tolerance model is the first opinion diffusion model to propose a truly dynamic threshold (i.e., a node's state evolves according to the dynamic interaction patterns). Therefore, based on its novelty and realism potential, we are encouraged to use the tolerance model in our paper.

2.1. The Tolerance-Based Opinion Diffusion Model. The tolerance model [45] is based on the classic voter model [49], being a refinement of the stubborn agent model [11, 52], with the unique addition of a *dynamic* decision-making threshold, called tolerance θ_i, for each node.

We further introduce the specific network science notations to mathematically define our model. Given a social network $G = \{V, E\}$, the neighbourhood of node $v_i \in V$ is defined as $N_i = \{v_j \mid (e_{ij}) \in E\}$. Exemplifying for a context with two competing opinions, we introduce two disjoint sets of stubborn agents $V_0, V_1 \in V$ which act as opinion sources. Stubborn agents never change their opinion, while all other (regular) agents $V \setminus \{V_0 \cup V\}_1$ update their opinion based on the opinion of one or more of their direct neighbours. We represent with $x_i(t)$ the opinion of agent v_i at time t. Normal (regular) agents start with a random opinion value $x_i(0) \in [0, 1]$. We represent with $s_i(t)$ the state of an agent v_i at moment t having continuous opinion $x_i(t)$. In case of a discrete opinion, representation $x_i(t) = s_i(t)$, and in case of a continuous opinion, representation $s_i(t)$ is given in the following equation.

$$s_i(t) = \begin{cases} 0 & \text{if } 0 \le x_i(t) < 0.5, \\ \text{none} & \text{if } x_i(t) = 0.5, \\ 1 & \text{if } 0.5 < x_i(t) \le 1. \end{cases} \tag{1}$$

In the assumed social network, agents v_i and v_j are neighbouring nodes if there is an edge e_{ij} that connects them. Some agents may not have an opinion or may not participate in the diffusion process (i.e., $s_i(t) = \text{none}$), so interacting with these agents will generate no opinion update. A regular node will periodically poll one random neighbour (simple diffusion) or all its neighbours (complex diffusion), average the surrounding opinion $\bar{x}_{N_i}(t)$ (i.e., vicinity N_i of an arbitrary node v_i, at time point t), and update its opinion $x_i(t)$ using a weighted combination of the past opinion and that of its neighbour(s), as

$$x_i(t) = \theta_i \cdot \bar{x}_{N_i}(t) + (1 - \theta_i) \cdot x_i(t - 1). \tag{2}$$

Alternate opinion assigning approach to balance propagation odds for both ranking methods (orangeorange & blue)
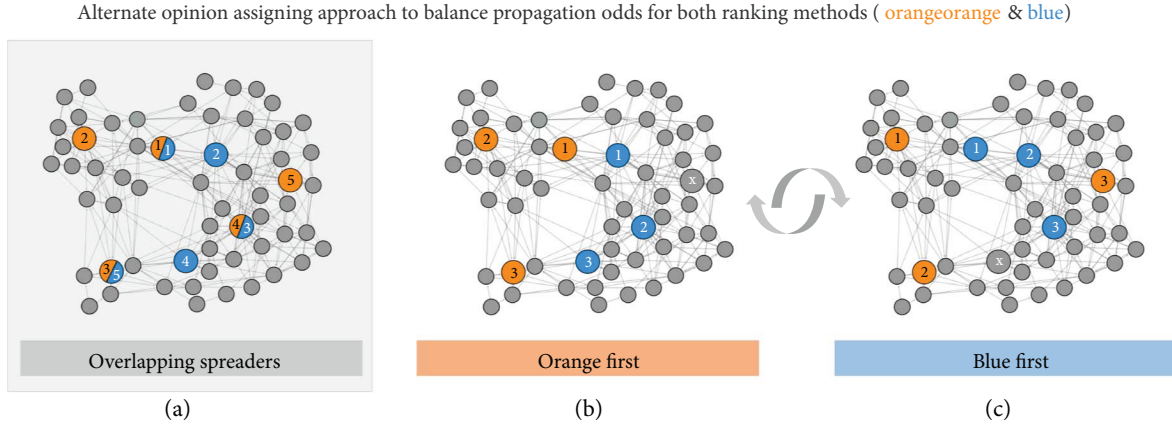

FIGURE 2: Example of the alternate opinion assigning approach in order to offer both competing ranking methods even chances of propagation. The coloured nodes marked with indices 1–5 represent the top 5 orange, respectively, blue spreaders, as determined by the two ranking methods. Moreover, some of these spreaders overlap ((a) e.g., 3/5 means 3rd best orange spreader and 5th best blue spreader), so we assign each spreader node one of two opinions (orange/blue) alternatively, starting with orange first (b) then blue first (c). As such, a simulation of *orange* versus *blue* ranking methods translates into two independent simulations, slightly favouring each method in turn. The assigning of opinion is always evenly distributed in terms of number of nodes, for example, 3 spreaders in this example.

The tolerance θ_i parameter is the amount of accepted external opinion and changes after each interaction based on whether a node has faced competing opinion or supporting opinion (in a binary context with opinions A and B). Once a node is in contact with the same opinion for a long enough time, it becomes intolerant ($\theta_i(t) = 0$), so that the network converges towards a state of balance [53]. Opinion fluctuates and is transacted by all nodes, but stubborn agents are the only nodes which do not become influenced in turn, acting as perpetual sources for the same opinion [11].

The evolution towards both tolerance and intolerance varies in a nonlinear fashion, as an agent under constant influence becomes indoctrinated at an increased rate over time. If that agent faces an opposing opinion, he will eventually start to progressively build confidence in the other opinion. As such, the tolerance model employs a nonlinear fluctuation function, unlike most models in literature [54, 55]. Based on realistic sociopsychological considerations in the dynamical opinion interaction model, we model tolerance evolution as

$$\theta_i(t) = \begin{cases} \max\left(\theta_i(t-1) - \alpha_0\varepsilon_0, 0\right) & \text{if } s_i(t-1) = s_j(t), \\ \min\left(\theta_i(t-1) + \alpha_1\varepsilon_1, 1\right), & \text{otherwise.} \end{cases}$$

$$(3)$$

Tolerance is decreased by $-\alpha_0\varepsilon_0$ if the state of the agent before interaction, $s_i(t-1)$, is the same as the state of the randomly interacting neighbour $s_j(t)$. If the states are not identical (i.e., opposite opinion), then the tolerance will be increased with the dynamic product of $+\alpha_1\varepsilon_1$. The two scaling factors, α_0 and α_1, both initialized with 1, act as weights (i.e., counters) which are increased to account for every event in which the initiating agent keeps its old opinion (i.e., tolerance decreasing) or changes its old opinion (i.e., tolerance increasing). Therefore, scaling factor α_0 is increased by +1 as long as an agent interacts with another agents having the same state (i.e., $s_i(t-1) = s_j(t)$) and is reset to 1 otherwise.

Scaling factor α_1 is increased as long as the interacting state is always different from that of the agent and is reset if the states are identical. We introduced the scaling factors to model bias and used to increase the magnitude of the two tolerance modification ratios ε_0 (intolerance modifier weight) and ε_1 (tolerance modifier weight). The two ratios are chosen with the fixed values of $\varepsilon_0 = 0.002$ and $\varepsilon_1 = 0.01$. We have determined these values as explained in [45].

In accordance with this presented mechanism, we designate two sets of stubborn agents, V_a and V_b, to act as initial spreaders *simultaneously*. In other words, we let all chosen centrality metrics compete against each other in a one-to-one diffusion scenario, where sets V_a and V_b consist of the top $p\%$ spreaders selected by each two pairs of centralities. We ensure that $V_a \cap V_b = 0$ and $|V_a| = |V_b|$, with $p = 0.05$. We find this approach to offer a good qualitative comparison basis for estimating the effectiveness of node ranking methods.

2.2. Alternate Opinion Assigning Approach. We further find that most state-of-the-art ranking methods have various degrees of overlapping in terms of the top spreader nodes they assign. As such, we introduce an alternate opinion assigning (AOA) approach in order to distribute nodes in the two sets of spreaders V_a and V_b evenly and equitable for both ranking methods, say A and B. Figure 2 exemplifies the AOA approach, where ranking method A is depicted with orange and method B is depicted with blue.

AOA means that each one-to-one influence ranking benchmark consists of two (or multiple of two) independent simulations. Considering that ranking methods A and B produce two partially overlapping sets of top $p\%$ spreaders, we alternate the simulations as follows:

(i) In the first simulation, method A *(orange)* has priority: one starts by assigning the first (top 1) spreader from V_a as an *orange* stubborn agent. This implies

TABLE 1: Graph statistics of the eight datasets detailing the number of nodes, edges, average degree ($\langle k \rangle$), maximum degree (k_{max}), average path length (APL), average clustering coefficient (ACC), and network diameter (Dmt).

Dataset	Nodes	Edges	$\langle k \rangle$	k_{max}	APL	ACC	Dmt
Random (Rand)	10,000	50,122	5.012	26	3.944	0.002	7
Mesh	10,000	53,896	5.39	44	11.51	0.148	30
Small-world (SW)	10,000	39,998	3.99	13	6.738	0.298	12
Scale-free (SF)	10,000	52,260	5.226	102	5.316	0.679	14
OSN	1899	20,296	10.68	339	3.055	0.138	8
Facebook (FB)	3172	94,458	29.78	470	3.714	0.501	10
LGU-emails (Emails)	12,625	20,362	3.226	576	3.811	0.577	9
POK	28,876	115,324	7.98	4305	4.05	0.076	13

that the spreader remains in V_a and is removed from V_b, if present.

(ii) Then, the first spreader from V_b is assigned as a *blue* stubborn agent, removing it from V_a, if present.

(iii) Alternatively, we assign nodes alternative opinion and filter them out from the other list of spreaders.

(iv) The AOA stops when $\min(|V_a|, |V_b|) = p \times N/2$ and discards any extra node so that $|V_a| = |V_b|$, ensuring that both sets V_a and V_b have an equal number of stubborn agents, namely, half of the desired $p \times N$ spreader population.

(v) In the second simulation, method *B (blue)* has priority: one starts by assigning the first (top 1) spreader from V_b as a *blue* stubborn agent. This implies that the spreader remains in V_b and is removed from V_a, if present.

(vi) The exact same AOA process is repeated, with B having priority over A.

The impact of AOA is highlighted in Figure 2, as we end up assigning two significantly different spreader sets for methods A and B. Methodologically speaking, one benchmark must consist of at least two simulations, but for better experimental results, one may run $2k$ simulations, ensuring that AOA is applied (i.e., k simulations favouring method A and k simulations favouring method B).

3. Results

We set out to discover fundamental drivers in the underlying graph structure which shape and influence opinion spreading in complex networks. To this end, our experimental setup is focused on a comparative benchmark analysis involving the reviewed node centrality metrics defined in Section 5.2. For an objective comparison, we make use of two types of datasets: synthetic data (10,000 node random, mesh, small-world, and scale-free networks [56]) and real-world data (consisting of large, representative complex networks sized between 1900 and 29,000 nodes).

In this section, two sets of results are detailed. First, we explore the correlations between ranking methods for assigning top spreaders. Naturally, within the top $p\%$ of nodes ranked by different centralities, we will eventually find common nodes. As such, we detect the amount of node overlapping $O_{ab} = V_a \cap V_b$ and express the correlation of the two measures as $\text{corr}_{ab} = |O_{ab}|/|V_a|$ and $\text{corr}_{ab} \in [0, 1]$. For the second experimental phase of benchmarking influence ranking methods, we ensure that $V_a \cap V_b = 0$ by alternatively assigning a node to each set, while removing it from the list of candidates of the other centrality, as explained by the AOA approach (Figure 2).

3.1. Correlations between Influence Ranking Methods. Real-world datasets can be viewed as topological compositions of the basic graph properties found in synthetic Erdos-Renyi random (Rand), Forest-fire mesh (Mesh), Watts-Strogatz small-world (SW), and Barabasi-Albert scale-free (SF) networks [56–58], so we solely rely on measurements on the synthetic datasets from Table 1. As such, the correlation process is applied on the four synthetic network types in order to better highlight distinguishable characteristic topological features, like uniform node degree distribution (random networks), high local clustering and community formation (mesh networks), and high clustering and long-range links (small-world), respectively, low average path length, and hub formation (scale-free).

Figure 3 presents the correlations corr_{ab} between 10×10 selected pairs of centralities; correlations are measured by considering the following spreader set sizes: $|V_a| = |V_b| = p \times N$, where $p \in \{0.01, 0.05, 0.1\}$ and N is the size of the graph, and find that corr_{ab} will drop slightly as p increases. The average changes δ in spreader correlations from $p = 0.01$ up to $p = 0.1$ are $\delta_{\text{Rand}} = -0.289$, $\delta_{\text{Mesh}} = -0.193$, $\delta_{\text{SW}} = -0.189$, and $\delta_{\text{SF}} = -0.088$. This overall drop in correlation can be explained as follows: more of the same nodes are determined as top spreaders by ranking methods when the spreader sets are small. As p increases, each ranking method adds more nodes to the set of spreaders and the chances of overlapping drop. However, when we look at each individual centrality measure in turn, we notice that some increase the correlation amount, while others drop that amount. Section 1 and Figure 1 in the Supplementary Materials detail and discuss these measurements for 10 selected ranking methods, over the four synthetic topologies, as p increases.

Rand	Deg	Cls	Btw	HITS	PR	HI	LR	KS	CLC	EC	*Avg*
Deg	–	0.480	0.803	0.976	0.968	0.576	0.976	0.128	0.536	0.552	0.665
Cls		–	0.608	0.488	0.464	0.552	0.472	0.036	0.872	0.856	0.536
Btw			–	0.824	0.808	0.600	0.808	0.052	0.632	0.624	0.639
HITS				–	0.976	0.622	0.976	0.092	0.544	0.552	0.669
PR					–	0.616	0.976	0.132	0.520	0.536	0.666
HI						–	0.618	0.060	0.560	0.536	0.524
LR							–	0.132	0.528	0.544	0.669
KS								–	0.144	0.136	0.101
CLC									–	0.952	0.587
EC										–	0.587

Mesh	Deg	Cls	Btw	HITS	PR	HI	LR	KS	CLC	EC	*Avg*
Deg	–	0.158	0.382	0.882	0.796	0.740	0.890	0.304	0.686	0.648	0.609
Cls		–	0.328	0.178	0.120	0.178	0.156	0.208	0.200	0.190	0.190
Btw			–	0.418	0.336	0.366	0.386	0.234	0.362	0.318	0.347
HITS				–	0.814	0.780	0.928	0.308	0.704	0.652	0.629
PR					–	0.628	0.874	0.222	0.558	0.516	0.540
HI						–	0.752	0.350	0.784	0.746	0.591
LR							–	0.278	0.664	0.620	0.616
KS								–	0.412	0.424	0.304
CLC									–	0.886	0.584
EC										–	0.555

SW	Deg	Cls	Btw	HITS	PR	HI	LR	KS	CLC	EC	*Avg*
Deg	–	0.356	0.450	0.530	0.528	0.506	0.534	0.190	0.422	0.424	0.437
Cls		–	0.730	0.612	0.602	0.602	0.604	0.122	0.796	0.778	0.578
Btw			–	0.836	0.826	0.798	0.828	0.118	0.862	0.794	0.693
HITS				–	0.990	0.934	0.992	0.110	0.786	0.750	0.726
PR					–	0.928	0.994	0.110	0.776	0.740	0.721
HI						–	0.930	0.112	0.766	0.740	0.701
LR							–	0.110	0.778	0.742	0.723
KS								–	0.112	0.116	0.122
CLC									–	0.902	0.688
EC										–	0.665

SF	Deg	Cls	Btw	HITS	PR	HI	LR	KS	CLC	EC	*Avg*
Deg	–	0.746	0.783	0.937	0.921	0.878	0.937	0.349	0.825	0.847	0.820
Cls		–	0.730	0.762	0.746	0.788	0.762	0.349	0.873	0.852	0.734
Btw			–	0.810	0.836	0.751	0.810	0.349	0.746	0.751	0.729
HITS				–	0.958	0.894	1.0	0.349	0.847	0.873	0.825
PR					–	0.868	0.958	0.349	0.820	0.847	0.811
HI						–	0.894	0.360	0.878	0.926	0.804
LR							–	0.349	0.847	0.873	0.825
KS								–	0.185	0.196	0.315
CLC									–	0.942	0.773
EC										–	0.789

FIGURE 3: Ratio of nodes overlapping in the top 10% ($N = 10K$ nodes) of spreader assignment by 10 centrality metrics (degree, closeness, betweenness, HITS, PageRank, Hirsch index, LeaderRank, k-shell, local centrality, and eigenvector centrality) in an Erdos-Renyi (Rand) random network, mesh network, small-world (SW), and scale-free (SF) network. The *blue* colour intensity of a cell corresponds to the strength of correlation found in the symmetric cell, that is, $\text{cell}_{\text{colour}}(i, j) \sim \text{cell}_{\text{value}}(j, i)$. A stronger blue intensity denotes a stronger correlation.

As a representative overview, we present in Figure 3 only the results for $p = 0.1$. For each centrality combination, we provide the numerical correlation and a symmetric graphical correlation. For example, the correlation degree-Hirsch index in the random network is $\text{corr}_{\text{Deg-HI}} = 0.576$, which translates into a mid-blue gradient in the table symmetric cell HI-Deg. The last column in the table expresses the average correlation on each line. Summing up and averaging the values on the last column, we obtain the cumulated correlations for each topology as $\text{corr}_{\text{Rand}} = 0.552$, $\text{corr}_{\text{Mesh}} = 0.497$, $\text{corr}_{\text{SW}} = 0.606$, and $\text{corr}_{\text{SF}} = 0.741$.

Quantitatively and also intuitively, the highest spreader correlation is obtained on the scale-free network, as it naturally consists of a very small core of hub nodes. These hubs act like an invariant to p in the topology and are likely to be selected as top spreaders by all centrality measures. Even if p is changed, the correlation remains high (see Supplementary Materials, Section 1). On the opposite spectrum lie the

random and mesh topologies. Both are characterized by uniformity in node properties, so that various centralities will have a higher heterogeneity in their top spreader selection, leading to the smaller measured correlations. Lastly, the small-world network borrows the uniformity of meshes and the long-range links of a random network. Here, we measure a relatively high average correlation of 0.606, denoting that this network has a stable core of influential nodes, like the scale-free network.

Analysing each centrality in turn, we notice that there are higher correlations between ranking methods of the same category, for example, diffusion-based HITS, PageRank, and LeaderRank. Furthermore, some centralities are more suitable for some topologies and less efficient for others. For example, we confirm that degree is considerably more relevant for scale-free networks (correlation of 0.802 with other centralities), but only marginally relevant for the small-world network (correlation of 0.437). The same observation is consistent with closeness and betweenness. To better highlight the spatial overlapping of spreader nodes, we provide a visual example in the Supplementary Materials, Section 2.

Arching over the presented results, we motivate the usage of alternate opinion assigning (AOA), because we find high node overlapping, ranging between 30% and 70%, between all state-of-the-art centralities.

3.2. Independent SIR Simulations. For a comparative basis, we first estimate the efficiency of an influence ranking method by employing classic SIR simulation [41, 42]. In this sense, we measure both the *time* needed to infect the majority of nodes (expressed in simulation iterations τ) and the final *coverage* of the infection (expressed as a percentage ρ of the total network size). We use the following SIR-specific parameter values [40, 41]: $p = 0.05$ (i.e., top 5% nodes selected as spreaders), $k = 0.95$ (i.e., at least 95% population to be infected as a stop condition), $\lambda = 0.05$ (i.e., 5% probability to become infected during an interaction), and $\delta = 10$ (i.e., 10 iteration duration of infectious state for a node).

The simulation results in Table 2 represent averaged values for τ and ρ by running 10 repeated simulations on each dataset, for each individual ranking method (i.e., amassing to a total of $10 \cdot 8 \cdot 10 = 800$ simulations). Through these results, we want to highlight that running a diffusion process for each ranking method in an *individual* manner (i.e., one by one), the provided feedback regarding ranking efficiency, is often limited.

The results for most topologies are very close in terms of measured τ and ρ, suggesting that differentiation between ranking methods is unreliable. For instance, analysing the coverages ρ in Table 2, the average coverage for Rand is $\overline{\rho}_{\text{Rand}} = 95.47\%$ with a standard deviation of only $\sigma_{\text{Rand}} = 0.082$. The measured difference Δ between the most efficient ranking method (Hirsch index) and least efficient ranking method (degree) is only $\Delta_{\text{Rand}} = 0.3\%$ on the Rand network. Similarly, the standard deviations σ for real-world networks are $\sigma_{\text{OSN}} = 0.214$, $\sigma_{\text{FB}} = 0.042$, $\sigma_{\text{Emails}} = 0.230$, and $\sigma_{\text{POK}} = 0.273$. The differences Δ between the most and least efficient ranking methods are roughly $\Delta_{\text{OSN}} = 1.4\%$, $\Delta_{\text{FB}} = 0.4\%$, $\Delta_{\text{Emails}} = 2\%$, and $\Delta_{\text{POK}} = 5.5\%$. For a visual representation

TABLE 2: Performance of ranking methods expressed as the time τ needed to infest a network (lower is better) and the final coverage ρ, expressed as a percentage of the network size (higher is better), using SIR benchmarking.

	Rand	Mesh	SW	SF	OSN	FB	Emails	POK
Time τ								
Deg	30.1	40.7	115.3	69.2	31.4	24.7	36.1	64.6
Cls	30.4	54.7	116.2	78.2	32.5	25.1	43.7	66.8
Btw	30.3	51.1	116.4	65.2	32.8	23.9	35.2	62.2
HITS	30.4	43.0	118.2	68.0	32.8	24.5	35.8	61.5
PR	30.2	39.7	117.0	71.9	33.2	26.7	35.3	68.5
HI	30.8	47.1	115.0	69.0	31.8	25.1	34.2	66.8
LR	30.1	41.7	119.1	73.4	33.0	24.6	37.0	62.5
KS	30.1	52.3	117.6	71.0	33.6	26.6	36.9	64.5
CLC	30.4	47.9	121.2	71.2	31.2	26.5	37.0	63.2
EC	30.1	49.2	123.1	73.0	32.6	24.9	36.6	64.9
Cov ρ								
Deg	95.31	95.11	80.83	48.77	79.59	95.31	46.92	57.49
Cls	95.42	95.06	80.72	49.58	79.96	95.17	47.42	57.33
Btw	95.51	95.07	80.92	48.91	79.58	95.29	46.91	57.91
HITS	95.54	95.08	81.90	47.95	79.55	95.31	46.86	57.31
PR	95.48	95.09	81.60	48.77	79.35	95.28	47.04	57.67
HI	95.61	95.09	81.75	48.47	79.17	95.28	46.68	57.98
LR	95.50	95.05	80.59	48.78	79.37	95.28	47.01	57.31
KS	95.53	95.05	81.12	48.41	79.58	95.23	46.67	57.68
CLC	95.43	95.11	80.64	49.05	79.60	95.29	46.69	57.79
EC	95.44	95.11	81.03	48.10	79.70	95.25	46.73	57.99

of the coverage ρ benchmark results refer to Supplementary Materials, Section 4.

We consider these simulation results to highlight an overall lack of perspective regarding which ranking method is better on a given topology. Likewise, the best ranking methods are not consistent across datasets. For instance, HITS turns out to be the most efficient ranking method on a SW, but the least efficient on a SF network; Deg is least efficient on Rand, 2nd on Mesh, 7th on SW, and 6th on SF, yet it comes 8th if we average all results; Btw is the 5th on OSN, 4th on FB, 5th on Emails, and 3rd on POK, and comes 3rd overall. This kind of inconsistency further supports our claims for an improved type of benchmarking methodology.

3.3. Competition-Based Simulations. We let each of the $n = 10$ selected centrality measures compete in a one-to-one scenario over the 4 synthetic and 4 real-world datasets. Every dataset comprises a total of $n \times (n - 1)/2 = 45$ pairs of simulations, translating into $2 \times 45 = 90$ individual simulations due to AOA. For statistical rigour, each experiment is repeated 10 times, consisting of a simulation batch of 20 simulations, leading to $45 \times 20 = 900$ simulations per dataset, amassing to an overall $8 \times 900 = 7200$ unique experiments. The large quantity of numerical results is available in the Supplementary Materials, Section 3 and Tables 1 and 2.

TABLE 3: Average performance of the 10 ranking methods on the 8 datasets. Performance is expressed as opinion coverage (%) obtained in the one-to-one opinion diffusion competitions with every other ranking method.

	Rand	Mesh	SW	SF	OSN	FB	Emails	POK
Deg	66.18	71.26	68.94	61.71	52.76	56.18	63.52	63.28
Cls	23.02	5.47	11.39	1.83	2.55	11.49	2.40	45.78
Btw	66.15	42.93	56.96	62.78	40.37	57.51	58.33	58.27
HITS	66.28	69.32	76.92	61.63	64.42	62.10	63.56	63.09
PR	77.16	65.35	71.93	55.74	41.08	55.99	63.55	63.94
HI	12.13	52.82	33.25	54.72	24.23	41.36	39.60	36.30
LR	76.95	67.57	66.72	61.53	64.39	68.06	63.97	66.87
KS	0.99	39.65	37.87	45.89	28.77	28.87	42.07	13.33
CLC	33.93	52.36	60.24	26.99	44.74	55.91	48.43	48.01
EC	23.12	32.96	39.43	43.09	62.83	44.54	51.49	32.27

Condensing the simulation results, we present in Table 3 the average performance of the 10 ranking methods on the 8 datasets. This performance is quantified as an average percentage of opinion coverage obtained from the one-to-one competition benchmarks (e.g., HITS obtains a coverage of 65.23% on the OSN dataset).

Similar to the state-of-the-art SIR epidemic benchmarking, our obtained results are easy to understand and offer the possibility of direct comparison between ranking methods on the same dataset. On the other hand, we notice two improvements by applying our methodology:

(1) There is much higher variation between measures on the same dataset. For example, on the FB dataset, we obtain Deg = 59.31% and Cls = 4.28%, which suggest an obvious performance difference. On the other hand, using SIR as benchmark, the coverages are $\rho_{Deg} = 95.31\%$ and $\rho_{Cls} = 95.17\%$.

(2) There is greater emergent granularity between measures on different datasets. For example, Cls turns out to be much less efficient on a SF topology (1.99%) than on a SW topology (18.37%).

Assessing the results in Table 3, we find an objective comparison of state-of-the-art ranking methods used in current social networks research. Figure 4 presents these cumulated performance indicators; the top three ranking methods, according to our original proposed methodology, are Leader-Rank (LR), HITS, and node degree (Deg).

The cumulated results in Figure 4 are based solely on the 8 datasets used throughout the paper. With more datasets used, the averaged performances will slightly differ. However, valuable insight is further offered by the visualization of performances on each dataset in turn; these results are detailed in the Supplementary Materials, Section 5.

Additionally, we provide a suggestive visual example of the opinion coverages at the end of a simulation, after balancing is attained [53] with our used tolerance diffusion model [45]. The Mesh topology is exemplified here because it offers

the most intuitive 2D spatial feedback after applying a force-directed layout. To this end, Figure 5 shows the coverage of competing centrality measures in three different scenarios:

(i) Two ranking methods with high overlapping and balanced outcome: Deg (orange) 56.70% and LR (blue) 43.30% (Figure 5(a)).

(ii) Two ranking methods with moderate overlapping and inefficient seed selection for one method (Btw): LR (orange) 74.26% and Btw (blue) 25.74% (Figure 5(b)).

(iii) Two ranking methods with low overlapping and extreme outcome: Cls (orange) 5.24% and HI (blue) 94.76% (Figure 5(c)).

The validation of our novel benchmarking methodology employs a standard strategy for the selection of multiple spreaders. After a review of the most recent advances in complex network analysis, we find that the method of simply selecting the top spreaders from the entire network is consistently found throughout literature [35, 37, 38, 59–62]. Nevertheless, there are several alternatives for selecting multiple spreaders which we detail in the Supplementary Materials, Section 6.

3.4. Comparison between Benchmarking Methods. To highlight the superior quantitative power of our competition-based benchmark, we aggregate the results in Table 4. Here, we measure the difference $\Delta_{min-max}$ between the most and least efficient ranking methods and the difference Δ_{1-2} between the top two ranking methods, for each dataset in turn. Seeking higher overall differences, we find that our proposed benchmarking methodology is more insightful, in general, than the classic SIR benchmark. As such, when measuring $\Delta_{min-max}$, individual SIR benchmarking only manages to produce differences of $\approx 0.06 - 1.59\%$ (1.14% on average) between ranking methods, while our proposed solution offers differences of $\approx 80 - 98\%$ (91% on average). When trying to discern between the top 2 ranking methods on a particular dataset, SIR manages to place them apart by only $\approx 0 - 1.07\%$ (0.31% on average), while our method manages to produce higher differences within $\approx 0.28 - 8.75\%$ (3.56% on average).

Another advantage of our proposed method is the overall uniformity obtained for the performances of each centrality across the 8 selected datasets. For instance, if LR and HITS result as the most efficient spreading methods on one topology, their performance is replicated with high confidence on the other topologies as well. When employing SIR benchmarking, the performances are not consistent across datasets. This aspect is suggested visually in Figure 6, where we highlight the most (LR) and least (Cls) efficient centralities, as they are ranked over the 8 datasets. It is easy to notice how LR is positioned in the top 3 and Cls in the last 2-3 methods overall. In the individual SIR benchmarking, there is no such uniformity.

In conclusion, our benchmarking methodology—which is specifically designed for the *competitive* social network
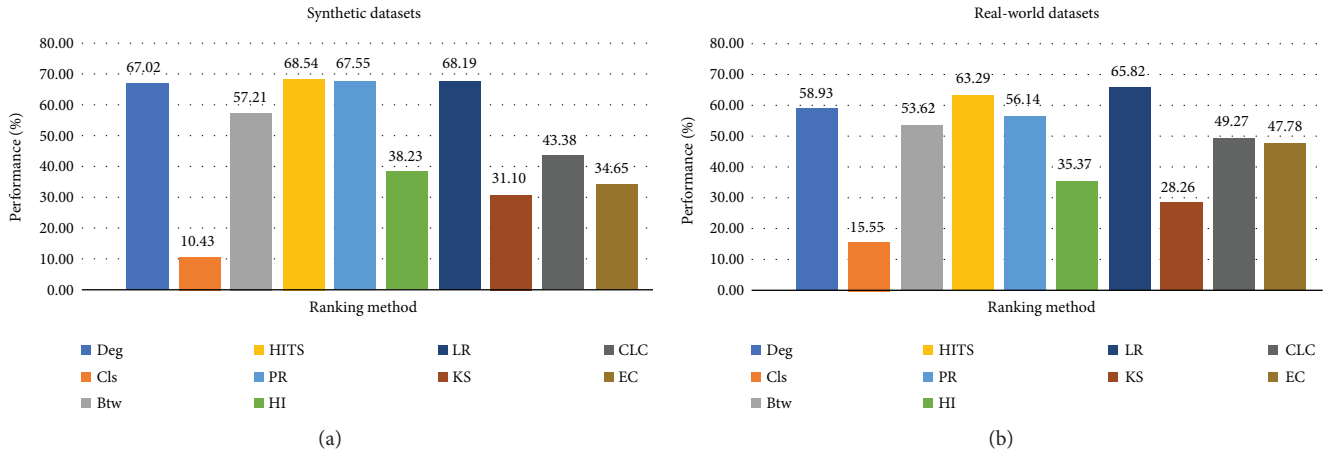
FIGURE 4: Coverage performance (0–100%) of each ranking method cumulated over all synthetic, respectively, all real-world datasets.
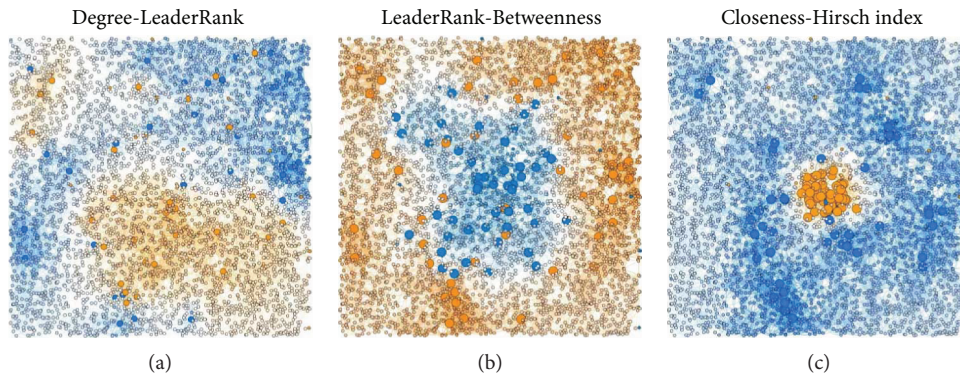

FIGURE 5: Three opinion diffusion benchmarks highlighting the final opinion coverage over the Mesh dataset ($N = 10,000$). Orange nodes are influenced more by the first ranking method, and blue nodes are influenced more by the second ranking method; whiter nodes are closer to indecision (50%); larger nodes represent seeders (1% of N).

TABLE 4: Comparison between individual SIR and our simultaneous competition-based benchmark in terms of how well ranking methods are differentiated. $\Delta_{min-max}$ is the difference (%) between the most and least efficient ranking methods; Δ_{1-2} is the difference (%) between the top 2 ranking methods on each dataset. Higher differences are better.

	(%)	Rand	Mesh	SW	SF	OSN	FB	Emails	POK
SIR	$\Delta_{min-max}$	0.31	0.06	1.59	3.28	0.99	0.15	1.58	1.17
Competition	$\Delta_{min-max}$	98.72	92.32	85.19	97.09	96.05	83.12	96.25	80.07
SIR	Δ_{1-2}	0.07	0.00	0.18	1.07	0.33	0.00	0.80	0.02
Competition	Δ_{1-2}	0.28	2.73	6.5	1.72	2.46	8.76	0.71	5.37

context—provides significant quantitative separation between influence ranking methods on synthetic and real social network topologies. This numerical separation is over one order of magnitude greater than the one provided by classic SIR simulation—a standard methodology used in epidemic spreading, where the diffusion context is less competitive and more ego-centred. Therefore, we encourage the use of our proposed method in specific real-world applications of dynamic social networks.

4. Discussion

One of the significant research challenges in network science is to rank a node's ability to spread information in a network [43]. As spreading is used to model real-world processes such as epidemic contagion and information propagation [2, 3, 20, 22, 63], our paper aims to improve current methodology in validating and comparing state-of-the-art ranking methods in the social network context. Numerous alternative ranking

Individual benchmarking								Competition-based benchmarking								
Random	Mesh	SW	SF	OSN	FB	Emails	POK	Random	Mesh	SW	SF	OSN	FB	Emails	POK	
1 HI	EC	HITS	Cls	Cls	Deg	Cls	EC	PR	Deg	HITS	Btw	HITS	LR	LR	LR	1
2 HITS	CLC	HI	CLC	EC	HITS	PR	HI	LR	HITS	PR	Deg	LR	HITS	HITS	PR	2
3 KS	Deg	PR	Btw	CLC	CLC	LR	Btw	HITS	LR	Deg	HITS	EC	Btw	PR	Deg	3
4 Btw	HI	KS	LR	Deg	Btw	Deg	CLC	Deg	PR	LR	LR	Deg	Deg	Deg	HITS	4
5 LR	PR	EC	PR	Btw	LR	Btw	KS	Btw	HI	CLC	PR	CLC	PR	Btw	Btw	5
6 PR	HITS	Btw	Deg	KS	PR	HITS	PR	CLC	CLC	Btw	HI	PR	CLC	EC	CLC	6
7 EC	Btw	Deg	HI	HITS	HI	EC	Deg	EC	Btw	EC	KS	Btw	EC	CLC	Cls	7
8 CLC	Cls	Cls	KS	LR	EC	CLC	Cls	Cls	KS	KS	EC	KS	HI	KS	HI	8
9 Cls	KS	CLC	EC	PR	KS	HI	LR	HI	EC	HI	CLC	HI	KS	HI	EC	9
10 Deg	LR	LR	HITS	HI	Cls	KS	HITS	KS	Cls	Cls	Cls	Cls	Cls	Cls	KS	10

FIGURE 6: Visual representation of the uniformity in benchmarking influence ranking methods across different networks. We highlight the positions obtained by LR (top centrality in terms of spreading) and Cls (least effective centrality) across our 8 datasets in the context of individual (a) and competition-based (b) benchmarks. The position of a centrality on the vertical corresponds to its obtained rank (1–10) after benchmarking. For example, LR is the 5th best on random and 10th best on Mesh.

methods have been developed, relying on classic graph centralities, localized targets [63], optimal percolation [43], and so on. While the challenge at hand remains partially unsolved, it is argued that insights are uncovered only through the optimal collective interplay of all the influencers in a network [43]. This emergent behaviour is also the key to our study, namely, the introduction of a benchmarking technique employing simultaneous competition-based spreading.

The main motivation of this paper is the need for increased realism in the social network context, where real-world applications imply simultaneous diffusion by their nature. Nevertheless, our methodology may be tailored to other interdisciplinary fields of science. One area of research that can benefit directly from our methodology is network biology. Specifically, determining node centrality is a hot topic in biological networks. For instance, a study shows that the phenotypic consequence of a single gene deletion is determined by the topological position in the molecular interaction network [64]; also, the relationship between the network roles of disease genes and their tolerance to germs shows that cancer driver genes occupy the most central positions [65]. Many biological studies rely on the theoretical results from network science, and they often only employ degree and betweenness centrality in their analysis. With our study, we aim to broaden the methodological perspective for interdisciplinary fields.

We find advantages over existing benchmarking methodology relying on the SIR epidemic model. Notably, our competition-based method offers much greater quantitative separation between ranking methods on the same dataset (e.g., degree is roughly 14 times more performant than closeness on the Facebook dataset); also, we obtain higher granularity for a ranking method on different datasets (e.g., closeness is roughly 9 times less efficient on a scale-free topology than on a small-world topology).

Further development ideas of our method are possible. For instance, one can increase the number of spreaders acting simultaneously in a network from 2 to $k > 2$. Accordingly, alternate opinion assigning (AOA) must be modified to fit the k opinion sources. The recent study discusses the importance of targeting specific localized targets, rather than obtaining a high coverage of the network [63]. Our method can be easily implemented to measure the target coverage during or at the end of a spreading simulation. Another study

finds that each complex network may have a small "control set" of nodes, which, when triggered, will influence the whole network [66]. These control sets are believed to be surprisingly small (5–10% of nodes) and may also be paired with our benchmarking methodology.

Finally, we consider that the topology-aggregated competition-based results we obtained (e.g., in Figure 4 of the Supplementary Materials) can be used to define a functional fingerprint of real-world networks based on how influence ranking methods perform on them. Namely, we notice that the 10 used centrality measures perform in a unique, distinguishable manner on the four fundamental synthetic topology models. This uniqueness can be quantified as a characteristic vector for random, mesh, small-world, and scale-free networks. Any real-world dataset can then be compared to other datasets through these four fingerprint vectors. Overall, we believe that our work improves a significant challenge in the study of opinion spreading phenomena and also serves as a good starting point for many of the still unsolved problems and new ideas found in literature.

5. Methods

5.1. Validation Datasets. We motivate the inclusion of *synthetic* datasets into the study to clearly distinguish between characteristic topological features of the network that influences spreading. These features include a normal versus power-law degree distribution, lower versus higher clustering, lower versus normal path lengths, existence of long-range links, or hub formation, respectively. The four chosen network models represent the four fundamental topology types out of which empirical networks are further built [26, 56, 57].

With a higher interest on influence spreading pertaining to the field of social network analysis, we choose four undirected (weighted and unweighted) networks consisting of various types of social relationships. As such, we rely on a weighted online social network (OSN) with 1899 users [67], an unweighted Facebook friendship network (FB) consisting of the 3172 students from a Computer Science faculty in Romania [68], an unweighted email exchange network (Emails) from London's Global University with 12,625 contacts [69], and a weighted friendship network (POK) with 28,876 users from the Slovakian POK platform [70]. On the other hand, all synthetic networks consist of 10,000 nodes

and are algorithmically generated using default parameters found in the state of the art. Table 1 provides the basic statistics for each such network.

5.2. Influence Ranking Methods. In order to define each centrality metric, we make use of the following graph theory-specific notations. A social network is a graph $G = (V, E)$ formed out of $|V|$ number of nodes and $|E|$ number of edges. The edges may also be directed (i.e., $e_{ij} \neq e_{ji}$) or weighted (i.e., they have weights w_{ij}). The connectivity of the graph is characterized by an adjacency matrix $\mathbf{A} = \{a_{ij}\}$, where $a_{ij} = 1$ (or w_{ij} in weighted context) if nodes v_i and v_j are connected and 0 otherwise. Furthermore, the degree of a node v_i is denoted as k_i, the neighbourhood of a node is the set of nodes $v_j \in N_i$, and the average degree of G is $\langle k \rangle = 2E/V$.

The reviewed measures considered for benchmarking in this paper are classified in one of three categories: structure-based, location-based, and diffusion-based rankings.

5.2.1. Structure-Based Measures. Structure-based measures require the topological information of the graph—either local (e.g., ego network, vicinity) or global (e.g., path-based). Under local measures, we first mention degree centrality (Deg) k_i of a node v_i; it is easy to use and efficient but less relevant in some real-world scenarios [34, 38], as some studies show that Deg fails to identify influential nodes because it is limited to the ego network of each node [34, 71].

The local centrality (LC) measure was introduced as a trade-off between the low-relevant degree centrality and other time-consuming measures [34]. LC of node v_i considers both the nearest and the next nearest neighbours and is defined as

$$
\begin{aligned}
\mathrm{LC}(v_i) &= \sum_{v_j \in N_i} Q(v_j), \\
Q(v_j) &= \sum_{v_k \in N_j} N(v_k),
\end{aligned}
\tag{4}
$$

where N_i is the vicinity (set of neighbours) of node v_i, $N(v_k)$ is the number of the nearest and the next nearest neighbours of node v_k, and $Q(v_j)$ is sum of $N(v_k)$ over each node in N_i. LC can be considered as more effective than degree centrality because it uses more information from the vicinity of distance 2 but has much lower computational complexity than betweenness and closeness centralities.

Another method considered a local ranking measure is ClusterRank (CR), proposed by Chen et al. [35]. CR quantifies the influence of a node v_i by taking into account not only its direct influence (out-degree k_i^{out}) and influences of its neighbours (like in the case of PageRank) but also its clustering coefficient c_i [56]. Formally, the ClusterRank score C R(v_i) of a node v_i is defined as

$$
\mathrm{CR}(v_i) = f(c_i) \sum_{v_j \in N_i} \left(k_i^{\text{out}} + 1 \right),
\tag{5}
$$

where the term $f(c_i)$ represents the effect of v_i's local clustering, the term +1 results from the contribution of v_j itself, and

N_i is the vicinity of node v_i. Based on empirical analysis [35], the authors propose the exponential function $f(c_i) = 10^{-c_i}$.

The local centrality with a coefficient, denoted as CLC by Zhao et al. [71], is a combination of the previous CR and LC methods. The number of neighbouring nodes is measured to identify cluster centres and is combined with a decreasing function f for the local clustering coefficient of nodes, called the coefficient of local centrality $c(v_i)$, namely, $f(c(v_i)) = e^{-c(v_i)}$. Mathematically, the influence of node v_i is measured as

$$
\mathrm{CLC}(v_i) = f(c(v_i)) \cdot \mathrm{LC}(v_i).
\tag{6}
$$

Considering the global information of the graph can give better insights, so we adopt the widely used betweenness (Btw) and closeness (Cls) centralities [56]. Betweenness of a node v_i is expressed as the fraction of shortest paths between node pairs that pass through the node v_i and is defined as [26]

$$
\mathrm{Btw}(v_i) = \sum_{i \neq j \neq k \in G} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}},
\tag{7}
$$

where σ_{jk} is the number of shortest paths between nodes v_j and v_k and $\sigma_{jk}(v_i)$ denotes the number of shortest paths between v_j and v_k which pass through node v_i.

Closeness centrality of a node v_i is defined as the inverse of the sum of distances to all other nodes in G; it can be considered as a measure of how long it will take to spread information from a given node to other reachable nodes in the network [56]:

$$
\mathrm{Cls}(v_i) = \left(\sum_{v_j \in G \setminus v_i} d(v_i, v_j) \right)^{-1}.
\tag{8}
$$

5.2.2. Location-Based Measures. Location-based measures also require the structural information of the graph but focus around the belief that the location of a node in a network is a more relevant. Driven by the limitations of simple graph metrics, such as degree centrality, Kitsak et al. propose k-core decomposition to quantify a node's influence based on the assumption that nodes in the same shell have similar influence, and nodes in higher-level shells are likely to infect more nodes [28]. To this end, the k-core decomposition method was validated by several studies [28, 29]. While this method is often found in literature under both the names of k-core or k-shell decomposition, the two concepts differ. The k-core of a graph is the maximal subgraph such that every vertex has degree at least k. A k-shell (KS), on the other hand, is the set of vertices that are part of the k-core but not part of the $(k + 1)$th-core.

Experiments show that by running a diffusion process on the network (e.g., SIR), the nodes with the same k_s values always have different number of infected nodes, namely, spreading influence [32]. This phenomenon suggests that the k-core decomposition method is not appropriate for ranking the global spreading influence of a network. Liu et al. [32] propose to solve this observed drawback by taking

into account the shortest distance between a target node and the node set with the highest k-core value. In terms of the distance from a target node v_i to the network core G_c, the spreading influences of the nodes with the same k-core values can be distinguished using the following equation:

$$\theta(v_i | k_s) = (k_s^{\max} - k_s + 1) \sum_{v_j \in G_c} d_{ij}, \quad i \in G_{k_s}. \tag{9}$$

In (9), $k_s^{\max}$ is the largest k-core value of G, d_{ij} is the shortest distance from node v_i to node $v_j \in G_c$, G_c is the network core, and G_{k_s} is the node set whose k-core values equal k_s.

In this paper, we also make use of the Hirsch index. The h-index (HI) [72] is a hybrid location-local-based centrality in which every node needs only a few pieces of information: the degrees of its neighbours. It was originally developed as a means to measure the scientific impact of scholars, but it now finds uses in quantifying the influence of users in social networks or drugs in pharmacological interaction maps. The h-index of a node v_i is defined as the largest value h so that v_i has at least h neighbours with a degree $\geq h$.

The algorithm is intuitive to apply, namely, for a node v_i with vicinity N_i, we order all its neighbours $v_j \in N_i$ in descending order of their degree k_{v_j}. The h-index $\mathrm{HI}(v_i)$ is the position $h - 1$ in the ordered list of nodes at which the degree of a neighbour becomes smaller than the position in the list. For example, given the list of degrees $L(v_i) = \{10, 8, 7, 6, 3, 1, 1\}$, we deduce $\mathrm{HI}(v_i) = 4$, because $L(v_i)[4] > 4$, but $L(v_i)[5] < 5$.

5.2.3. Diffusion-Based Measures. Diffusion-based measures are based on obtaining a state of balance in the network after applying a nondeterministic spreading processes, like a random walk. We make use of the fundamental eigenvector centrality (EC), which supposes that the influence of a node is not only determined by the number of its neighbours (i.e., degree centrality) but also by the influence of each neighbour [73]. Inspired by EC, there are three additional algorithms we discuss in this paper.

PageRank (PR) was first implemented as a random walk on the network of hyperlinks between web pages [74]. A damping factor d is introduced as the probability for a user to jump to a random website, and $1 - d$ is the probability for the user to continue browsing through hyperlinks. The influence $s_t(v_i)$ of a node v_i at time t is given by

$$\mathrm{PR}(v_i) = \frac{1-d}{|V|} + d\left(\sum_{v_j \in G} \frac{\mathrm{PR}(v_j)}{k_j^{\mathrm{out}}}\right), \tag{10}$$

where $|V|$ is the number of nodes in G, k_j^{out} is the out-degree of node v_j, and $d = 0.85$, but d requires step-wise optimization based on the network.

HITS is similar to PR, based on the concept that good hub nodes will point to good authority nodes, and good authorities will point by good hubs [75]. The hub score of

all nodes at time $t = 0$ is initialized with 1; the authority score $\mathrm{Aut}_t(v_i)$, at any moment in time t, is expressed as

$$\mathrm{Aut}_t(v_i) = \sum_{v_j \in G} a_{ji} \cdot \mathrm{Hub}_{t-1}(v_j),$$

$$\mathrm{Hub}_t(v_i) = \sum_{v_j \in G} a_{ji} \cdot \mathrm{Aut}_t(v_j). \tag{11}$$

Finally, the LeaderRank (LR) algorithm represents an improvement over PR, since the probability parameter is adaptive, leading to a parameter-free algorithm directly applicable on any type of the complex network [37]. The method is applied by adding an additional ground node v_g that is connected to all other nodes, ensuring the graph is connected. A random walk then adds a score of +1 to each visited node v_i. The ground node starts with $s_g(0) = 0$, and all other nodes in G have $s_i(0) = 1$. Using the notation $s_t(v_i)$ at time t for a node v_i, the evolving score can be expressed as

$$s_{t+1}(v_i) = \sum_{v_j \in G} p_{ij} s_t(v_j) = \sum_{v_j \in G} \frac{a_{ij}}{k_i^{\mathrm{out}}} s_t(v_j). \tag{12}$$

The score $s_t(v_i)$ is proven to converge towards a steady state at time t_c [37]; the score of the ground node is then evenly distributed to all other nodes $V \in G$ to conserve the scores on the nodes of interest. The final, stable LR score is expressed as

$$\mathrm{LR}(v_i) = s_{t_c}(v_i) + \frac{s_{t_c}(v_g)}{|V|}. \tag{13}$$

Data Availability

The real-world social network datasets supporting this research article are from previously reported studies and have been cited individually in the Methods section. The generated (synthetic) datasets are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

Supplementary Materials

Figure 1: changes in correlation of node overlapping, for the 10 analysed ranking methods, as the spreader size p is increased from 1% to 10% of the total network size N. Each synthetic network has $N = 10{,}000$ nodes. Figure 2: spatial distribution of selected spreader nodes on the mesh network with $N = 10{,}000$ nodes. The top $p = 1\%$ nodes are highlighted

as spreaders, as determined by the degree, closeness, betweenness, and PageRank centralities, respectively. Table 1: synthetic dataset (i.e., random, mesh, small-world, and scale-free) benchmark results for pair-wise competition between centrality measures. Each cell (x, y) contains the final opinion coverage (0–100%) for centrality x; the symmetric cell (y, x) represents the same number on a colour gradient blue (0%), white (50%), and orange (100%). Table 2: real-world dataset benchmark results for pair-wise competition between centrality measures. Each cell (x, y) contains the final opinion coverage (0–100%) for centrality x; the symmetric cell (y, x) represents the same number on a colour gradient blue (0%), white (50%), and orange (100%). Figure 3: performance of each ranking method (i.e., coverage 0–100%) on the 8 datasets using individual SIR benchmarking. Figure 4: performance of each ranking method (i.e., coverage 0–100%) on the 8 datasets using simultaneous competition-based benchmarking. Figure 5: comparison between the naïve (a–c) and graph colouring (d–f) methods using three competitive diffusion examples on the mesh network ($N = 10,000$ nodes). Larger nodes represent spreader nodes. The first centrality in the figure captions corresponds to orange opinion and the second centrality to blue opinion. Figure 6: difference in spreader spacing for closeness (orange) when switching from the naïve method (a) to the graph colouring method (b). Table 3: comparison between the naïve and graph colouring methods in terms of selecting spreader nodes. Performance is expressed as percentage (%) for each node centrality in three competitive simulation scenarios. *(Supplementary Materials)*

References

[1] S. Eubank, H. Guclu, V. S. Anil Kumar et al., "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.

[2] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabasi, "The human disease network," *Proceedings of the National Academy of Sciences of United States of America*, vol. 104, no. 21, pp. 8685–8690, 2007.

[3] S. B. Rosenthal, C. R. Twomey, A. T. Hartnett, H. S. Wu, and I. D. Couzin, "Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion," *Proceedings of the National Academy of Sciences of United States of America*, vol. 112, no. 15, pp. 4690–4695, 2015.

[4] B. Karrer and M. E. J. Newman, "Competing epidemics on complex networks," *Physical Review E*, vol. 84, no. 3, article 036106, 2011.

[5] B. A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos, "Winner takes all: competing viruses or ideas on fair-play networks," in *Proceedings of the 21st international conference on World Wide Web - WWW '12*, pp. 1037–1046, Lyon, France, 2012.

[6] A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos, "Interacting viruses in networks: can both survive?," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '12*, pp. 426–434, Beijing, China, 2012.

[7] F. Darabi Sahneh and C. Scoglio, "Competitive epidemic spreading over arbitrary multilayer networks," *Physical Review E*, vol. 89, no. 6, article 062817, 2014.

[8] L.-C. Chen and K. M. Carley, "The impact of countermeasure propagation on the prevalence of computer viruses," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 2, pp. 823–833, 2004.

[9] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '03*, pp. 137–146, Washington, DC, USA, 2003.

[10] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman, "E-mail as spectroscopy: automated discovery of community structure within organizations," *The Information Society*, vol. 21, no. 2, pp. 143–153, 2005.

[11] D. Acemoğlu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 1–27, 2013.

[12] M. Magdon-Ismail, M. Goldberg, W. Wallace, and D. Siebecker, "Locating hidden groups in communication networks using hidden Markov models," in *Intelligence and Security Informatics*, pp. 126–137, Springer, 2003.

[13] E. Adar and L. A. Adamic, "Tracking information epidemics in blogspace," in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, pp. 207–214, Compiegne, France, 2005.

[14] S. González-Bailón, J. Borge-Holthoefer, and Y. Moreno, "Broadcasters and hidden influentials in online protest diffusion," *American Behavioral Scientist*, vol. 57, no. 7, pp. 943–965, 2013.

[15] E. Ferrara, P. De Meo, S. Catanese, and G. Fiumara, "Detecting criminal organizations in mobile phone networks," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5733–5750, 2014.

[16] T. Ritter, I. F. Wilkinson, and W. J. Johnston, "Managing in complex business networks," *Industrial Marketing Management*, vol. 33, no. 3, pp. 175–183, 2004.

[17] R. Brennan, *Business-To-Business Marketing*, Springer, 2014.

[18] A. G. Haldane, "Rethinking the financial network," in *Fragile Stabilität – Stabile Fragilität*, pp. 243–278, Springer, 2013.

[19] D. Y. Kenett, T. Preis, G. Gur-Gershgoren, and E. Ben-Jacob, "Dependency network and node influence: application to the study of financial markets," *International Journal of Bifurcation and Chaos*, vol. 22, no. 07, p. 1250181, 2012.

[20] P. Csermely, T. Korcsmáros, H. J. M. Kiss, G. London, and R. Nussinov, "Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review," *Pharmacology & Therapeutics*, vol. 138, no. 3, pp. 333–408, 2013.

[21] L. Udrescu, L. Sbârcea, A. Topîrceanu et al., "Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing," *Scientific Reports*, vol. 6, no. 1, 2016.

[22] S. Pei and H. A. Makse, "Spreading dynamics in complex networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2013, no. 12, 2013.

[23] M. Li, H. Zhang, J.-x. Wang, and Y. Pan, "A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data," *BMC Systems Biology*, vol. 6, no. 1, p. 15, 2012.

[24] M. S. Granovetter, "The strength of weak ties," *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[25] J. Kleinberg, "Small-world phenomena and the dynamics of information," in *Advances in Neural Information Processing Systems*, pp. 431–438, Curran Associates, 2002.

[26] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[27] A.-L. Barabási, *Network Science*, Cambridge University Press, 2016.

[28] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[29] C. Castellano and R. Pastor-Satorras, "Competing activation mechanisms in epidemics on networks," *Scientific Reports*, vol. 2, no. 1, p. 371, 2012.

[30] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," *Physics Letters A*, vol. 377, no. 14, pp. 1031–1035, 2013.

[31] J.-H. Lin, Q. Guo, W.-Z. Dong, L.-Y. Tang, and J.-G. Liu, "Identifying the node spreading influence with largest *k*-core values," *Physics Letters A*, vol. 378, no. 45, pp. 3279–3284, 2014.

[32] J.-G. Liu, Z.-M. Ren, and Q. Guo, "Ranking the spreading influence in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 18, pp. 4154–4159, 2013.

[33] Y. Liu, M. Tang, T. Zhou, and Y. Do, "Improving the accuracy of the *k*-shell method by removing redundant links: from a perspective of spreading dynamics," *Scientific Reports*, vol. 5, no. 1, p. 13172, 2015.

[34] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 4, pp. 1777–1787, 2012.

[35] D.-B. Chen, H. Gao, L. Lü, and T. Zhou, "Identifying influential nodes in large-scale directed networks: the role of clustering," *PLoS One*, vol. 8, no. 10, article e77455, 2013.

[36] S. Pei, L. Muchnik, J. S. Andrade, Z. Zheng, and H. A. Makse, "Searching for superspreaders of information in real-world social media," *Scientific Reports*, vol. 4, no. 1, p. 5547, 2014.

[37] L. Lü, Y.-C. Zhang, C. H. Yeung, and T. Zhou, "Leaders in social networks, the delicious case," *PLoS One*, vol. 6, no. 6, article e21202, 2011.

[38] Q. Li, T. Zhou, L. Lü, and D. Chen, "Identifying influential spreaders by weighted LeaderRank," *Physica A: Statistical Mechanics and its Applications*, vol. 404, pp. 47–55, 2014.

[39] Z.-M. Ren, A. Zeng, D.-B. Chen, H. Liao, and J.-G. Liu, "Iterative resource allocation for ranking spreaders in complex networks," *EPL (Europhysics Letters)*, vol. 106, no. 4, p. 48005, 2014.

[40] M. E. J. Newman, "Spread of epidemic disease on networks," *Physical Review E*, vol. 66, no. 1, article 016128, 2002.

[41] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.

[42] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of Modern Physics*, vol. 87, no. 3, pp. 925–979, 2015.

[43] F. Morone and H. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, no. 7563, pp. 65–68, 2015.

[44] J. Cannarella and J. A. Spechler, "Epidemiological modeling of online social network dynamics," 2014, http://arxiv.org/abs/1401.4208.

[45] A. Topirceanu, M. Udrescu, M. Vladutiu, and R. Marculescu, "Tolerance-based interaction: a new model targeting opinion formation and diffusion in social networks," *PeerJ Computer Science*, vol. 2, article e42, 2016.

[46] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: a survey," *ACM SIGMOD Record*, vol. 42, no. 1, pp. 17–28, 2013.

[47] M. Granovetter, "Threshold models of collective behavior," *American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.

[48] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: a complex systems look at the underlying process of word-of-mouth," *Marketing Letters*, vol. 12, no. 3, pp. 211–223, 2001.

[49] R. A. Holley and T. M. Liggett, "Ergodic theorems for weakly interacting infinite systems and the voter model," *The Annals of Probability*, vol. 3, no. 4, pp. 643–663, 1975.

[50] R. Axelrod, "The dissemination of culture: a model with local convergence and global polarization," *Journal of Conflict Resolution*, vol. 41, no. 2, pp. 203–226, 1997.

[51] K. Sznajd-Weron and J. Sznajd, "Opinion evolution in closed community," *International Journal of Modern Physics C*, vol. 11, no. 6, pp. 1157–1165, 2000.

[52] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione, "Binary opinion dynamics with stubborn agents," *ACM Transactions on Economics and Computation*, vol. 1, no. 4, pp. 1–30, 2013.

[53] M. Udrescu and A. Topirceanu, "Probabilistic modeling of tolerance-based social network interaction," in *2016 Third European Network Intelligence Conference (ENIC)*, pp. 48–54, Wroclaw, Poland, 2016.

[54] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, 2002.

[55] W. Weidlich, "Sociodynamics––a systematic approach to mathematical modelling in the social sciences," *Chaos, Solitons & Fractals*, vol. 18, no. 3, pp. 431–437, 2003.

[56] X. F. Wang and G. Chen, "Complex networks: small-world, scale-free and beyond," *IEEE Circuits and Systems Magazine*, vol. 3, no. 1, pp. 6–20, 2003.

[57] A. Topirceanu, A. Duma, and M. Udrescu, "Uncovering the fingerprint of online social networks using a network motif based approach," *Computer Communications*, vol. 73, pp. 167–175, 2016.

[58] A. Topirceanu and M. Udrescu, "Statistical fidelity: a tool to quantify the similarity between multi-variable entities with application in complex networks," *International Journal of Computer Mathematics*, vol. 94, no. 9, pp. 1787–1805, 2017.

[59] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[60] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, "Breakdown of the internet under intentional attack," *Physical Review Letters*, vol. 86, no. 16, pp. 3682–3685, 2001.

[61] R. Pastor-Satorras and A. Vespignani, "Immunization of complex networks," *Physical Review E*, vol. 65, no. 3, article 036104, 2002.

[62] J.-G. Liu, J.-H. Lin, Q. Guo, and T. Zhou, "Locating influential nodes via dynamics-sensitive centrality," *Scientific Reports*, vol. 6, no. 1, article 21380, 2016.

[63] Y. Sun, L. Ma, A. Zeng, and W.-X. Wang, "Spreading to localized targets in complex networks," *Scientific Reports*, vol. 6, no. 1, article 38865, 2016.

[64] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41-42, 2001.

[65] J. Piñero, A. Berenstein, A. Gonzalez-Perez, A. Chernomoretz, and L. I. Furlong, "Uncovering disease mechanisms through network biology in the era of next generation sequencing," *Scientific Reports*, vol. 6, no. 1, article 24570, 2016.

[66] V. Nicosia, R. Criado, M. Romance, G. Russo, and V. Latora, "Controlling centrality in complex networks," *Scientific Reports*, vol. 2, no. 1, p. 218, 2012.

[67] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social Networks*, vol. 31, no. 2, pp. 155–163, 2009.

[68] A. Topirceanu and M. Udrescu, "ACSANet: ACSA social network dataset collection," 2014, http://cs.upt.ro/~alext/acsanet.

[69] H. Makse, "Software and data," http://www-levich.engr.ccny.cuny.edu/webpage/hmakse/software-and-data/.

[70] L. Takac and M. Zabovsky, "Data analysis in public social networks," in *International Scientific Conference and International Workshop Present Day Trends of Innovations*, pp. 1–6, Łomża, Poland, 2012.

[71] X. Zhao, F.'a. Liu, J. Wang, and T. Li, "Evaluating influential nodes in social networks by local centrality with a coefficient," *ISPRS International Journal of Geo-Information*, vol. 6, no. 2, p. 35, 2017.

[72] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 46, pp. 16569–16572, 2005.

[73] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Networks*, vol. 29, no. 4, pp. 555–564, 2007.

[74] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank Citation Ranking: Bringing Order to the Web*, Tech. Rep., Stanford InfoLab, 1999.

[75] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.