

Research Article

Towards Exploring the Influence of Community Structures on Information Dissemination in Sina Weibo Networks

Zhiwei Zhang , Aidong Fang , Lin Cui , Zhenggao Pan , Wanli Zhang ,
Chengfang Tan , and Chao Wang 

School of Informatics and Engineering, Suzhou University, Suzhou, China

Correspondence should be addressed to Lin Cui; jsjxcuilin@126.com

Received 5 May 2021; Accepted 4 August 2021; Published 14 August 2021

Academic Editor: Binxiang Dai

Copyright © 2021 Zhiwei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The power of online social networks to propagate information within communities and from one community to the next is undeniable. Both network structure and information propagation affect each other; they restrict and cooperate with each other. However, they can also dynamically reshape the network topology of user's social relationship in this process. The above process ultimately forms a feedback loop: the network structure affects how information spreads, while information propagation reshapes network topologies, so both evolve in concert over time. Using information propagation trees (IPT) of posts from the Sina Weibo microblogging site, we conducted a null model-based analysis to determine the influence of community structures on information propagation. We first generated randomized copies of the IPTs and then mined community structures from the originals and copies for comparison. An in-depth examination of the results in terms of improved significant profile, the length of information propagation path, and the relevance of the nodes in the propagation path indirectly reveals the inhibitory effect of community structures on information propagation.

1. Introduction

The way people gain and propagate information has undergone revolutionary changes with the advent of Web 2.0. Users are no longer just consumers of information; they are also its producers, disseminators, and critics [1, 2]. In no small part, social networks have shaped these new modes of communication. Arguably, social network platforms are now the mainstream of information publishing. However, beyond forming huge interpersonal networks of friends new and old where information is shared, social networks can burgeon information feedback mechanisms [3, 4], and a good recent review is information cascades in complex networks [5].

According to Yang et al. [6], the more people interact, the more information that is propagated, and the more information that is propagated, the more that people interact. As time goes by, this cyclical phenomenon propagates more and more information both dynamically and collaboratively. For the purpose of empirically studying the

influence of information propagation on network topologies and vice versa, we need to study the phenomenon caused by information dissemination and analyze the main factors that affect the propagation of information [7, 8]. Understanding that phenomena and factors might be the key to controlling rumor mongering on the web [9, 10]. This paper utilizes a null hypothesis-based IPT null model to research the laws of information propagation and the impact of community structures on information dissemination, further finding answers from an in-depth case analysis of Sina Weibo.

We all know the maxim “birds of a feather flock together.” This is concept sociologists use to describe how people with same or similar interests, hobbies, habits, values, and so on to interact and establish corresponding social relations. Studies show that the same applies to the virtual world [11, 12], and particularly to social networks. Online communities tend to revolve around particular values, which can reveal clues about their composition, structural features, functional properties, and dynamics in different granularities [13, 14]. For example, the online fan groups

(communities) of different film or movie stars in Sina Weibo represent the offline real groups formed based on same interests and common idol. Perception and prediction of information propagation in online social networks, including microprediction of individual's perspective on whether a person can disseminate information, medium prediction of discussion topic activity and propagation of local community groups, and macroprediction of information propagation breadth, depth, and speed in global perspective, play an important role in effectively understanding and developing online networks, which lay the foundation to further formulate corresponding control strategies to prevent rumors and other fake information dissemination, while releasing positive energy information for guidance [15, 16].

However, today's internet is far more complex than it used to be [5]. The applications and platforms that form online communities commonly span many disciplines, hence conducting research in those social networks increasingly requires interdisciplinary integration [17]. In terms of application, recent research has shown that influence has important application in social networks in [17]. Moreover, the lines between online and offline networks are blurring, network structures are becoming ever more dynamic, and single networks are merging to form multiple networks, mega networks, and multilayer monolithic networks. All these things only add to the challenges of the uncertainty of revealing the laws and patterns of information propagation and the difficulty of mining communities from the network and lack the corresponding comparative benchmarks for a specific online social network, such as Sina Weibo and Twitter, while the research on related social science issues is also more dynamic and systematic [15].

For these reasons, the null hypothesis in statistics is extended for analyzing the characteristics of Sina Weibo IPTs, which is highly likely that many of the findings in studies are from the Web 1.0 era, such as [15, 18, 19], no longer apply, and the laws and patterns of online information propagation have never been comprehensively mapped. In view of issues aforementioned, we studied the null model with different orders of IPTs in our previous work [20, 21] and explored a triad percolation method for community detection in [11], respectively.

Our objective with this paper is to comprehensively analyze how community structures can influence information propagation. To accomplish this goal, we conducted a null model analysis using Tsinghua University's Sina Weibo dataset (<https://github.com/garnettyige/datasets>), which contains information propagation trees (IPTs) that trace the real-world spread of information on topics from politics to fashion. To generate a null model, we generated a second dataset containing randomized copies of the Sina Weibo IPTs. Then, we mined community structures from the IPTs in both datasets and compared the results. The differences between the two in terms of significant profile should reveal the phenomenon of information propagation and show us how community structures affect information dissemination.

However, the main differences between the IPT null model and the traditional information propagation models is that the IPT null model can produce randomized copies of a real IPT, and the IPT randomized copies contribute to the benchmarks needed for comparative analysis. Besides, a traditional statistical analysis would not produce reliable results about the rules and laws of information propagation. Therefore, the analysis of IPT information dissemination based on the null model is more universal for the analysis of different topic information, especially for the dissemination analysis of network emergencies without reference benchmarks.

Thus, the three main contributions this research makes to the literature include

- (1) A null-model empirical analysis that reflects how community structures can influence information propagation
- (2) A null model that can generate a randomized copy of an IPT, and a community detection algorithm based on triad percolation that can discover the communities contained in both the original IPT and a randomized copy
- (3) A method for integrating triad percolation community detection and IPTs to conduct null model analyses

The remainder of this paper is organized as follows. Section 2 illustrates the research results and works related to the topic of this article. Section 3 introduces preliminary of the Sina Weibo IPTs, motivating force and ideas. Section 4 outlines the IPT null model and evaluation metrics. In Section 5, purpose experimental analysis is reported, and Section 6 offers concluding remarks and sheds light on future research directions.

2. Related Works

There has been much fruitful research on information propagation in social networks, with many models attempting to explain how information spreads and stops spreading through communities. Among these models, perhaps the most basic and common is the infectious disease model, which studied the disease transmission speed, spatial range, transmission route, and dynamic mechanism of infectious diseases, so as to guide the effective prevention and control of infectious diseases. Likewise, the influence of communities on information propagation has also been studied extensively, with tree-based models, including IPTs, at the fore of these efforts. Hence, the relevant literature on each of these models is briefly reviewed in this section.

2.1. How Communities Shape Information Propagation: The Infectious Disease Model. The most classical propagation model in the field of information propagation is the infectious disease model [18]. This model divides the population into different categories or partitions for different stages of the disease. Pastor-Satorras et al. [19] modeled information propagation in social networks by simulating

infection processes in complex networks, focusing on the main modes and outcomes of “disease” transmission. Their results have universal significance. Cannarella and Spechler [22] used the extended irSIR infectious disease model to explain in detail how information in Myspace and Facebook is taken up and discarded.

In response to observations that there was very little extensive research on the propagation of rumors in online social networks [22, 23], Lind et al. [24] proposed a basic rumor propagation model that introduces two variables: one indicating the maximum mean probability that two nodes will exchange a rumor with each other, and the other indicating the minimum propagation time required for the above exchange process to take place [25]. To investigate how inherent information propagation rates and the non-linearity of online social networks affect rumor dissemination in scale-free social networks, Roshani and Naimi [26] proposed a general SIS-based rumor propagation model. Myers and Leskovec [27] studied competition and cooperation between different types of information propagation in the real world and found that competitive propagation may reduce the probability that information will spread, while cooperative information dissemination positively influences whether a person will pass on information.

Although these infectious-disease-type models have had remarkable research consequences [24, 26–28], the results are mostly macroinsights, for example, the number of susceptible populations, the number of groups infected, and the quantity of recovered. The factors that intrinsically influence the spread of information have not yet been studied in depth.

2.2. How Information Propagation Reveals Communities: Tree-Based Models. The propagation paths left by information in online social networks can be used to reconstruct network communities. A typical way of representing these paths is with IPTs. Yi et al. [29] redefined the concept of event outbreaks in social networks based on IPTs. With them, they mined the key predictors of event outbreaks in social networks, with research results that have significantly improved the accuracy and timeliness of microblog event outbreak prediction. Fabrega and Paredes [30] analyzed the “three degrees of influence” phenomenon on Twitter, revealing that more than 90% of microblog posts only spread through three or fewer degrees of a network. However, further research by Rodrigues et al. [31] shows that the spread width of the URL propagation tree in Twitter is significantly greater than its depth; moreover, the spread width about 0.1% of IPT is larger than 20, while the height about only 0.005% of the IPT is greater than 20. Thus, the phenomenon of high width and low depth appears, which is in sharp contrast to the phenomenon of narrow width and high depth mentioned in [32]. In response to the findings of Rodrigues et al. [31], Golub and Jackson [33] conducted a study in the structural characteristics of online IPTs, and they found that different strategies could fundamentally change how information was disseminated via e-mail—a question first raised by Liben-Nowell and Kleinberg [32].

2.3. Null Models and Network Analysis. As is well known, in statistics, a network G with property P , such as average degree and clustering coefficient, also has a certain property Q , which is the null hypothesis. To verify the null hypothesis, it is necessary to construct randomized copies of the network with the same number of nodes and the same property P as benchmarks for comparison collectively referred to as the null model. Then, if the null model also has the property Q , Q must be a representative feature of this type of network.

Null models are widely used by academic researchers. For example, Maslov and Sneppen [34] applied a null model to study the specificity, stability, and topological characteristics of protein interaction and gene regulation networks. Maslov et al. [35] proposed a general scheme for analyzing and discovering topological patterns in large-scale complex networks that involves measuring the correlations between neighboring nodes. With their method, they proved that the characteristics of a community in a network closely relate to the degree distribution and correlation profile of the network.

Without loss of generality to topological characteristics, we use a null model to generate randomized copies of the Sina Weibo IPTs. Then, by mining the communities in the original IPTs and the copies and comparing the results in terms of significant profile, propagation distance and burstiness, we can determine the laws of information dissemination and how community structures affect information dissemination.

3. Preliminaries and Motivations

3.1. Information Propagation Tree. Consider a microblogger who posts a message that is then forwarded by his friends and friends of friends, and so on. In the representation of an IPT, as shown in Figure 1, the user is the tree’s root node (yellow), the friends are the green and gray nodes in the next layer down; the friends of friends are nodes in a layer below that, and so on, and the directed arrows indicate the propagation paths. The gray nodes signal that the information was passed on. A comprehensive description of the process for constructing IPTs is provided in [29] and the authors’ previous works [20, 21].

3.2. Null-Model Analysis. We have two overall goals with this research. The first is to gain an in-depth understanding of the factors that promote or inhibit the spread of information through communities. And, the second is to identify whether how the spread of information shapes communities in turn. To explore these questions, we chose a null model analysis for three key reasons.

- (1) Null models are an important type of the stochastic model that can be used to compare a specified model with the randomized theoretical model. To determine the role of a variable or mechanism, it is necessary to determine whether the model itself or the index correlated to the variable is significantly different from the null model and its related parameters [33]. Our variable is node connection

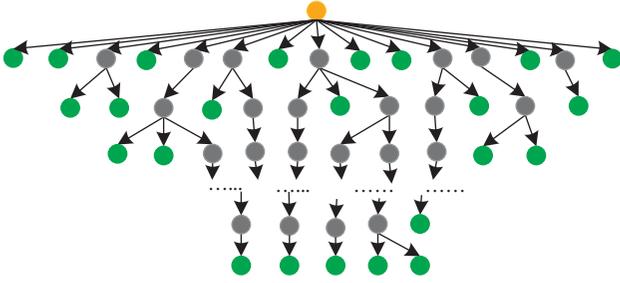


FIGURE 1: Example of an IPT. The yellow root node is the user who posted the original message. Gray nodes are users who spread the information (propagation nodes) and green nodes indicate users who did not spread the information (terminal nodes) [20]. The directed arrows show information flows. Hence, an IPT is also a subgraph of the social networks formed by users, i.e., one of the subgraphs of the user’s circle of friends.

probability, and we want to confirm that the community structure in the IPT has an impact on information diffusion by fully applying the advantages of a null model test.

- (2) The propagation of information or the occurrence of events is accidental. Due to the accident, the events seem to be correlated with each other, such as the information forwarding between users. However, it is still considered that their relationship is irrelevant in statistics, and so it is still appropriate to describe their relationship with the null hypothesis in statistics. Fortunately, the null model is based on the null hypothesis of occasionality in statistics. Thus, the correlations between events are reliable and considerable.
- (3) In practical terms, IPTs can only ever show a small part of how and where some information spread. The Sina microblog IPTs contain millions of nodes and edges, but still, this is almost certainly just a fraction of these bloggers’ social networks and whom they may have shared the information with. Hence, a traditional statistical analysis would not produce reliable results about the rules and laws of information propagation. However, by analyzing randomized copies of IPTs generated with a null model, we can considerably generalize the data to improve the reliability of the results.

3.3. Datasets. For our empirical analyses and tests, we used IPTs from the Sina microblog prepared by Tsinghua University (<https://github.com/garnettyige/datasets>). The dataset contains 340 IPTs, each fully tracing the dissemination of one piece of information. In total, the 340 IPTs involve 4,469,809 nodes (users) and 4,469,469 directed edges (propagation paths). The topics of discussion cover social issues, economy, sports, technology, fashion, health, and culture. Further details about the dataset are available at [29].

3.4. Motivations. Although the information propagation in online social networks, such as Sina Weibo, Twitter, and Facebook, has been extensively explored, existing research

studies more often than not focus on statistical analysis based on the network structure itself and lack a unified and scalable information dissemination analysis model. An IPT of Sina Weibo indicates the real and complete path of a microblog information in the Sina Weibo microblogging site, and we apply the null model to IPTs of Sina Weibo and then analyze the structure characteristics of IPTs, including community structures, so as to realize the analysis of information propagation indirectly. Moreover, the null model is the premise of IPTs’ analysis, which is extensible and can be served as a random benchmark for comparative analysis; it also has the advantage of computational complexity in the process of verification and comparison between the proposed algorithm and random variation method.

4. Methods and Metrics

This section sets out the analysis methods and settings used in the analysis. The key to constructing a good null model for IPTs is the edge rewiring strategy, which begins the section. We then discuss the community detection process and conclude with the metrics used to evaluate the results.

4.1. Null Model Edge Rewiring Strategy. This section aims at the influence of the community structure in online social networks on information propagation, combined with authors’ previous research results and construction process of Sina Weibo IPTs’ T [20]; the nodes in the lower layer of the randomized copy T_r of T acquire information from the nodes in the upper adjacent layer with a certain probability P , and the nodes degree in T follows the power-law distribution [20]. As a result, the edge rewiring strategy adopted in this paper is the extension and application of Price [36, 37], and the lower layer nodes in two adjacent layers of T_r obtain information from the upper layer nodes with probability P , that is, each node in the lower layer selects a node from the upper adjacent layer with probability P as the starting point of information propagation directed edge.

In prior work related to this research, the authors have designed three kinds of null models for IPTs based on the correlations between nodes, cascading rates, and the burstiness of information propagation. The three models are a statistical constrained 0-order null model, a random-rewire-broken-edges 0-order null model, and a random-rewire-broken-edges 2-order null model. More details on each of these models can be found in [20].

In this paper, we combine the edge rewiring strategy illustrated in our previous work [20], redefining the edge connection probability P of T_r , as per the following equation:

$$P_i = \frac{\left(\vec{k}_i + 1\right)}{\sum_{j=1}^{N_{\text{level}}} \left(\vec{k}_j + a\right)}, \quad (1)$$

where \vec{k}_i represents the out-degree of node i in $level$ -th layer of T_r , N_{level} indicates the number of nodes in $level$ -th layer, and a is a constant used to regulate the preferential choice mechanism and the connection probability between nodes.

In terms of the nodes in two adjacent layers, node j in the lower layer selects node i from the upper adjacent layer with probability P_i as the starting point of information propagation with node j as the endpoint, as indicated by the directed edge $i \rightarrow j$. Moreover, the randomized copy T_r of T established by P has the characteristic of community as explained in [20, 36, 37].

Just as some posts “go viral,” others go nowhere—not one person forwards the information. While authentic, analyzing information that does not spread at all is not useful for our purposes. Therefore, to avoid this situation occurring in a random copy, we set the value of a to 1. Thus, equation (1) can be simplified as follows:

$$P_i = \frac{\binom{\vec{k}_i + 1}{}}{\sum_{j=1}^{N_{\text{level}}} \binom{\vec{k}_j + 1}{}}, \quad (2)$$

$$P_i = \frac{1}{N_{\text{level}}}. \quad (3)$$

Intuitively, equation (2) can be further reduced to equation (3) for the initial state of the IPT, which is exactly consistent with the power-law distribution of the degree of nodes in social networks. Hence, it is guaranteed that the randomized copy of the IPT will carry the same community characteristics as the original with this edge rewiring strategy. The algorithm for generating the randomized IPTs according to this process.

4.2. Community Detection in IPTs. As is common knowledge, online social networks in the real world are sparse, and the degree of nodes follows the power-law distribution. Therefore, in reality, it is not wise to rely on there being a k -clique [38] or a largest clique [39]. However, each node must be contained in an open triad or a closed triad, without considering single free nodes or single edges in the network. For simplicity, in this paper, all nodes are located in open triads, and all communities were found in the following manner: (1) all open triads are extracted from the IPT; (2) using a triad percolation method similar to CPM [38], an initial community is formed by expanding a selected open triad as an initial seed; (3) an IPT community was formed by iteratively expanding the initial community until no community could meet the criteria for being deemed a community detailed in [20]; (4) the above process is repeated until there are no open triads. Figure 2 illustrates the full procedure.

4.3. Evaluation Metrics. We use two metrics in our empirical analysis: (1) cascade rate (CRP) along the information propagation path to measure the information that escapes a community; (2) significant profile measures the statistical significance of the difference between a specific length/cascade rate that occurs in the original IPT compared to its copies. Each is explained in more detail below.

As a measurement for null model comparison, significant profile measures the degree correlation between two

connected nodes. However, when analyzing the impact of communities on information propagation, it is not only necessary to consider the information dissemination within a community but also the information dissemination between the communities, which the degree correlation cannot measure.

To extend the traditional significant profile for this purpose, we introduce the cascade rate (CRP) along the information propagation path to measure the information that escapes a community. CRP explains the relationship between the degree correlation of an information propagation path (IPP) and the length of those paths. CRPs are calculated as follows:

$$\text{CRP}(j) = \frac{\sum_{i \in \text{IPP}} \vec{k}_i}{\sum_{n \in N} \vec{k}_n}, \quad (4)$$

where N is the set of nodes contributing to the total CRP of an information propagation path of length j . After all, we take the mean value $\text{SP}(j)$ of the M IPPs with the same length of j as the final cascade rate of the IPPs with a length of j , where M represents the total number of IPPs with length j in an IPT, and $\text{SP}(j)$ is defined in the following equation:

$$\text{SP}(j) = \sum_{j=1}^M \frac{\text{CRP}_i(j)}{M}. \quad (5)$$

However, $\text{SP}(j)$ not only explains the correlation between the length of the information propagation path and the nodes but also reflects the depth of the information propagation. Correspondingly, the R -value R_V of the cascading rate of the information propagation paths is defined as follows [20]:

$$R_V(j) = \frac{\text{SP}(j)}{\text{SP}_r(j)}, \quad (6)$$

where $\text{SP}_r(j)$ is the counterpart of $\text{SP}(j)$ in the randomized copy of an IPT. In addition, when $R_V(j) > 1$, the length j of the information propagation path is positively correlated with the degree correlation of the nodes, and negatively correlated otherwise.

Without loss of generality, we further measure the statistical significance of the difference between the frequency with which an information propagation path of a specific length/cascade rate occurs in the original IPT compared to its copies. For this metric, we employ the statistics Z -test and the Z -value definition of the Z -test, as shown in the following equations [20]:

$$Z_V(j) = \frac{\text{SP}(j) - \text{SP}_r(j)}{\sigma_{r_p}(j)}, \quad (7)$$

$$Z_P_j = \frac{Z_V(j)}{\left(\sum Z_V(i)^2\right)^{1/2}}, \quad (8)$$

where $\sigma_{r_p}(j)$ is the standard deviation of $\text{SP}_r(j)$ of the information propagation path of length j in the copy and $Z_V(\cdot)$ is the Z -value of all information propagation paths in an IPT.

Input: T , an IPT.

Output: T_r , the corresponding randomized copy of T .

- (1) **Initialization:** Starting from the root node of T , the nodes in each layer of T are copied layer by layer to T_r .
- (2) Acquire the node set N of each layer of T_r , $N = \{N_1, N_2, \dots, N_n\}$, where N_i ($1 \leq i \leq n$) stands for the node set of i th layer in T_r .
- (3) **for** each $N_i \in N$ **do**
- (4) **for** each node $v \in N_{i+1}$ **do**
- (5) $m = A$ node is randomly selected from N_i with the probability $P_i = (1/|N_i|)$ in equation (3).
- (6) A directed edge $m \rightarrow v$ is established in T_r .
- (7) $N_i = N_i - m$.
- (8) **end for**
- (9) **end for**
- (10) **return** T_r .

ALGORITHM 1: NMRC-IPT: null model-based algorithm for generating the randomized copy of an information propagation tree. The algorithm will generate a randomized copy T_r of the IPT T according to the edge rewiring strategy detailed in equation (3).

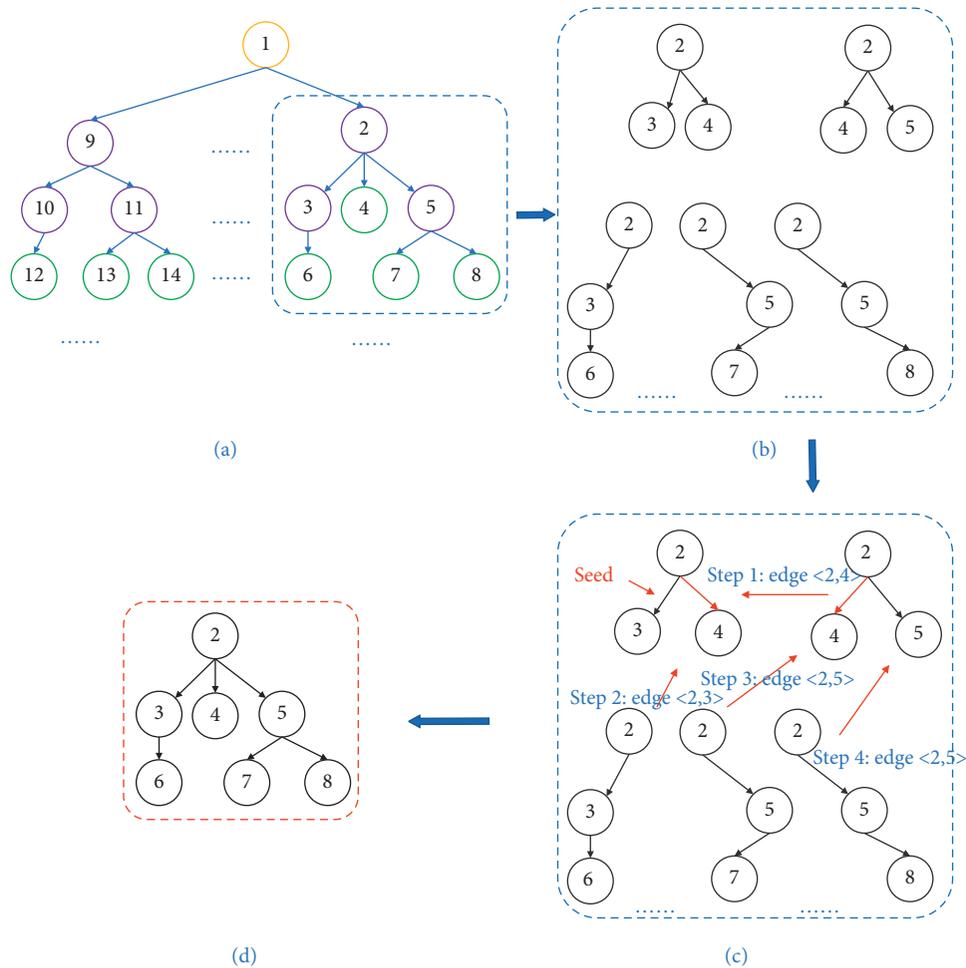


FIGURE 2: Community detection procedure in an information propagation tree with the triad percolation. (a) An example IPT; (b) open triads extracted from the IPT; (c) the procedure: first, an open triad is selected randomly as the initial seed for community C . Then, an open triad ot is selected and merged into C if ot and C have the same edge, and a new community C' comes into being. Finally, $C = C'$. The above process is repeated until the community meets the evaluation criteria detailed in [11], or there are no more open triads need to be merged; (d) the final community partition.

The larger the absolute value of the Z -value, the greater the difference and vice versa. Again, we extended the traditional significant profile by adding the R -value and Z -value of degree

correlation as well as the R -value and Z -value of the CRP on the same plane figure, which we call as extended significant profile (ESP). Algorithm 2 summarizes the metric calculations.

Input: T , an IPT.

Output: the extended significant profile (ESP) of T .

(1) The randomized copy T_r of T is generated by NMRC-IPT detailed in Algorithm 1.

(2) Community detection in T and T_r by utilizing the method detailed in Section 4.2. And, the community partition of T and T_r is $CT = \{ct_1, ct_2, \dots, ct_m\}$ and $CR = \{cr_1, cr_2, \dots, cr_n\}$, respectively.

(3) **for** each $ct_i \in CT$ **do**

(4) The R -value R_V of ct_i is calculated by equation (6).

(5) $R_{V_t} = R_{V_t} \cup R_V$.

(6) The Z -value Z_V of ct_i is calculated by equation (8).

(7) $Z_{V_t} = Z_{V_t} \cup Z_V$.

(8) **end for**

(9) **for** each $cr_i \in CR$ **do**

(10) The R -value R_V of cr_i is calculated by equation (6).

(11) $R_{V_r} = R_{V_r} \cup R_V$.

(12) The Z -value Z_V of cr_i is calculated by equation (8).

(13) $Z_{V_r} = Z_{V_r} \cup Z_V$.

(14) **end for**

(15) The mean value of R_{V_t} , Z_{V_t} , R_{V_r} and Z_{V_r} is calculated as $R_{V_{ta}} = (R_{V_t}/|CT|)$, $Z_{V_{ta}} = (Z_{V_t}/|CT|)$, $R_{V_{ra}} = (R_{V_r}/|CR|)$, and $Z_{V_{ra}} = (Z_{V_r}/|CR|)$, respectively.

(16) The R -value and Z -value of degree correlation and the R -value and Z -value of the cascade rate of IPPs are also calculated, and plotting them on a same plan figure together with $R_{V_{ta}}$, $Z_{V_{ta}}$, $R_{V_{ra}}$, and $Z_{V_{ra}}$, as shown in Figure 3, i.e., the extended significant profile.

(17) **return** the ESP calculated in 16.

ALGORITHM 2: Null model of IPT-based information propagation analysis. Firstly, a randomized copy of the IPT is generated; then, the corresponding R -value and Z -value in the extended significant profile are calculated by equations (6)–(8); finally, the obtained R -value and Z -value are drawn on the same plane, so as to acquire the ESP for analyzing the laws and patterns of information propagation.

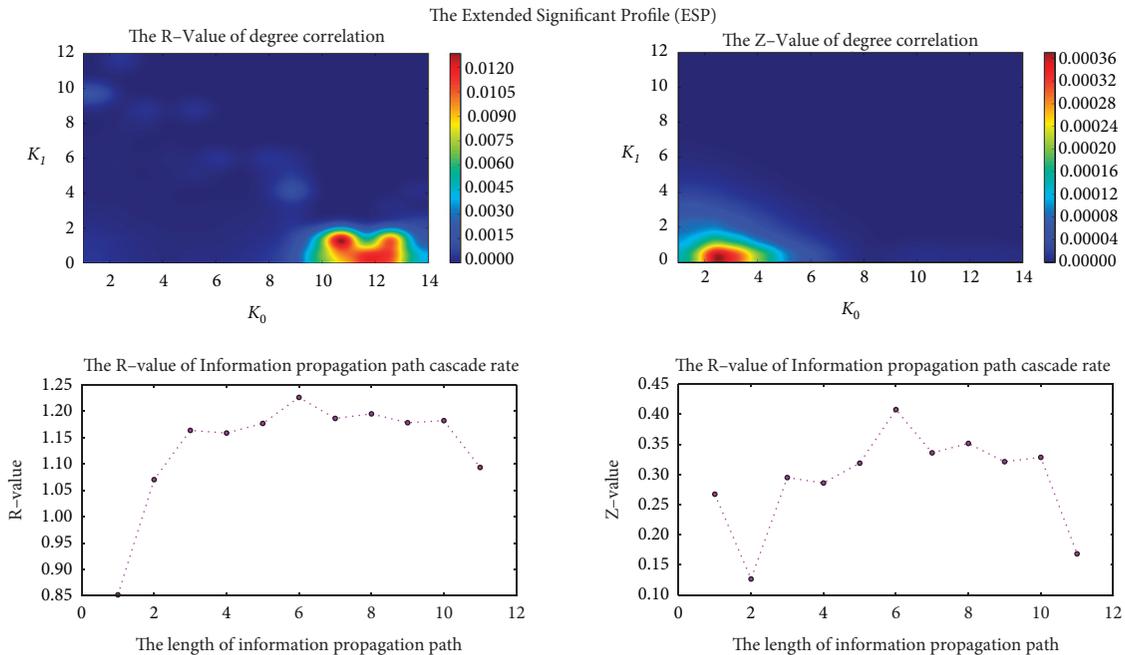


FIGURE 3: An example of extended significant profile [20]. This figure shows one example of an extended significant profile, i.e., the R -value and Z -value of degree correlation and the R -value and Z -value of the CRPs, which can be employed for analyzing the laws and patterns of information propagation.

4.4. Prior Results. This paper extends the results of prior work to explore the information dissemination mode and how community structures influence on information propagation in the Sina Weibo microblogging site. The basic

analyses of propagation width, propagation depth, degree distribution, degree correlation, R -value and Z -value of degree correlation, and R -value and Z -value of CPRs can be found in [20, 21]. These tests show that the network

topological structure played a decisive role in the information dissemination, and both the R -value and the Z -value of degree correlation are far less 1.0, and the connection between nodes with small out-degree $\in [1, 5]$ are strong restrained, while the connection between nodes with moderate out-degree $\in [7, 10]$ are weak restrained, and such weak connections also exist in nodes with bigger out-degree $\in ([10, \infty)$. The degree of nodes reflects the local aggregation of the network, which is the basis for the formation of community structures. All of the above have a direct guidance on the community detection and information dissemination analysis in this paper.

In the next section, we will conduct empirical analysis on Sina Weibo IPTs in terms of community partition modularity and explore how communities affect the information dissemination.

5. Empirical Analyses and Discussion

Our analysis begins with community characteristics analysis so as to gain the information spreading mode in Sina Weibo IPTs employed in Section 3.3. Then, we examine the impact of communities and their topic on information propagation.

5.1. Community Characteristics. Community structures are one of the topological characteristics of online social networks. As a subgraph of a social network, IPTs, therefore, reflect some of the characteristics of the community it represents. Figure 4(a) shows an example community structure from a Sina Weibo IPT. Note that the nested hierarchical structure depicted only has open triads [11], and open triads are formed between nodes (social network users), while there are no closed triads because the forward propagation of information does not establish backtracking paths, i.e., after users push messages to their friends, they will not receive the same messages from their friends.

The closer its value is to 1, the stronger the community structure divided by the network, that is, the better the partition quality:

$$CC = \frac{3 * \Delta}{(\Delta + \wedge)}. \quad (9)$$

The clustering coefficient is used to describe the degree to which nodes connected to the same center node in the network are also adjacent to each other. It is used to describe the probability that a person's friends in a social network are also friends with each other, reflecting the degree of nodes clustering in the network. According to the clustering coefficient CC in equation (9), where Δ represents the closed triad and \wedge stands for the open triad, we can see that the clustering coefficient of each node and the whole IPT is 0. This is a clearly unique property of IPTs as a subgraph of the user social relationship network of Sina Weibo. However, the above clustering coefficient does not fully explain all the characteristics of the community in the IPT. Hence, to detect the communities in the IPT, we used the technique detailed in [11, 40], we set the resolution coefficient to 1.0, which is consistent with literature [11, 40], and the corresponding

modularity distribution of the community partitions of each IPT, as shown in Figure 4(b).

A novel community structure mode of Sina Weibo IPT is illustrated, as shown in Figure 4(b), which can also reveal a new pattern of information diffusion. The results indicate that the community modularity of each IPT has an even and linear distribution, and the community modularity measures whether the community partition is a relatively better result; the higher the modularity, the better the community partition. From this, we can draw the following conclusions about information propagation in the Sina Weibo: (1) at the macrolevel, information spreads linearly in sharp contrast to the traditional random patterns associated with information dissemination; (2) locally, information only spreads among a user's circle of friends so as to form a local community structure; (3) because, Sina microbloggers are "birds of a feather that flock together," their similar interests and preferences mean information tends to spread widely and rapidly. Saturation comes quickly within communities but spreads to new communities more slowly, where the process repeats until the information stop spreading. This combination of rapid within-community transmission but slow intercommunity transmission acts to impede the spread of information.

5.2. Community Influence on Information Propagation. Figure 5 shows the communities extracted from a representative IPT in Sina Weibo about the topic of fashion.

From the example, we can see that the nodes within a community are closely connected, while the connections to other nodes in the network are relatively loose. This arrangement of friends is typical of the birds-of-a-feather phenomenon. The users within a community interact frequently, and information is spread widely between them. However, the interaction between communities is relatively sparse, which makes intercommunity diffusion weak. From this, we find that community structures may actually inhibit the spread of information in the Sina Weibo microblogging site.

There are many users with a high degree, often sitting at the center of communities, who post messages frequently. Information propagation begins when a high-degree user forwards information to their friends on a large scale. This is 1-degree of diffusion, i.e., $j = 1$. As shown in Figure 5, this is a clustered community structure centered on node A. Information also spreads to the communities centered on nodes B and C, where the within-community propagation process is similar to that of community A. However, the information propagation between communities is relatively weak, which shows that the community hinders the information diffusion. This is consistent with the results of our empirical analysis in [20] on the propagation depth and the node out-degree distribution of the Sina Weibo IPTs.

The overlap nodes between communities in the IPT are often opinion leaders with a large degree and located in the center of a community. The edges between communities A, B, and C show that weak connections between the nodes with larger degrees play a role in promoting information

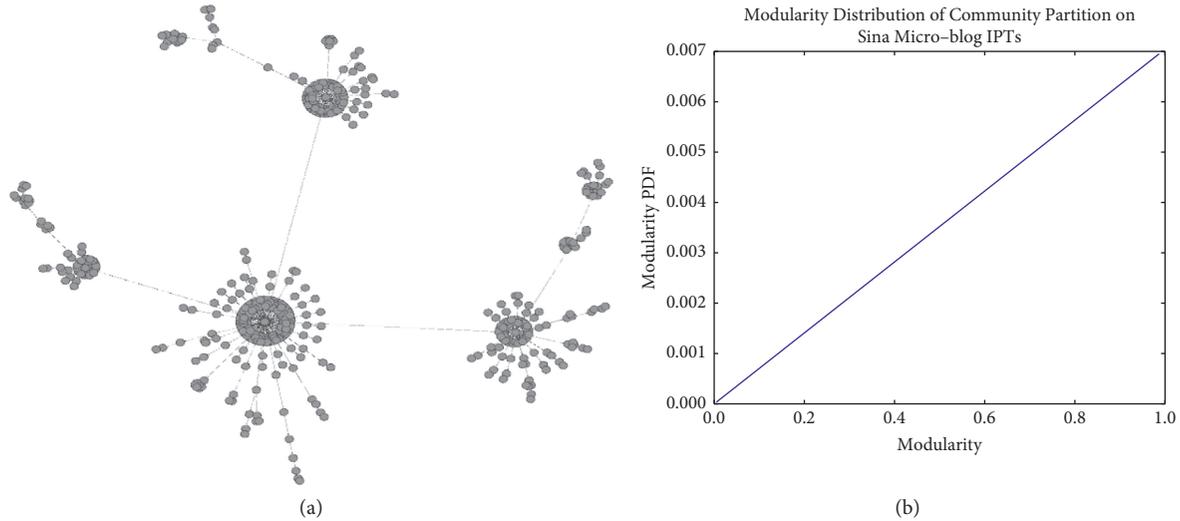


FIGURE 4: (a) An example of a community structure in one of the IPTs from the Sina Weibo; each cluster is a community structure, drawn by the Gephi; (b) the modularity distribution of community partition, which can be used to quantitatively measure the quality of network community partition.

propagation. Users at the community center are more likely to establish contact with other communities and spread information. When information spreads to another community, it is first propagated among the users in the community, which further inhibits information for far-reaching diffusion. The empirical and experimental analysis of the extended significant profile of Sina Weibo IPTs in [20], including degree correlation and the cascade rate of information propagation path, can prove that the connections between nodes with a large degree are suppressed, i.e., the information propagation is constrained. However, the connections between the nodes with a small degree are strong, which is caused by the sparse connection between the communities and the strong connections between the nodes within a community. IPTs with the above characteristics directly show that the community has a role in suppressing information diffusion and making information propagation paths shorter.

5.3. Topics and Information Propagation. Each message in an online social network has its own topic category. The IPTs used in Section 3.3 for empirical analysis and model verification cover topics spanning social events, economy, sports, technology, fashion, health, culture, and so on. IPT structures, however, are not dependent on topic. Rather, structural differences between IPTs are determined by the specific content in a post, the source user's social network, and the interactions between users. As an analysis model, the IPT null model proposed in this paper indirectly analyzes information propagation by analyzing the topological characteristics of the paths it took through a network. When we apply null model for analyze a specific topic of information propagation, it can completely ignore the topic properties of information itself and only consider the structural features. However, the general analysis steps are shown in Figure 6.

- (1) Assuming that there are m topic categories for the IPTs' dataset S ; then, $S = \{Z_1, Z_2, \dots, Z_m\}$, where Z_i ($1 \leq i \leq m$) represents the IPTs that belong to the topic of Z_i in S , and $Z_i = \{T_{i1}, T_{i2}, \dots, T_{in}\}$. For each IPT Z_{ij} ($1 \leq j \leq n$) that belongs to the topic Z_i , the corresponding randomized copy T_r is generated by the null model, and communities are detected from Z_{ij} and T_r , as shown in Figure 6.
- (2) Taking multiple T_r as the benchmarks, we compare and analyze the degree correlation, CRP , and community structure characteristics between Z_{ij} and T_r and take the mean value of multiple comparison results as the final correspondences to obtain the extended significant profile. Regarding the nodes with small out-degree of ($1 \leq K_1 \leq 2$) and large out-degree of ($9 \leq K_0 \leq 14$), the R -value of degree correlation is relatively larger, while the Z -value is relatively smaller, which illustrates that there is only one edge between those nodes.
- (3) Moreover, except for the R -value of information propagation path with length of 1 is less than 1, while for all other nodes is greater than 1, and the mean Z -value of the cascade rate of information propagation path is less than 30%, i.e., the mean error is less than 30%, which all not only elaborate that the out-degree distribution of nodes in Sina Weibo IPTs is consistent but also shows that the implementation of the information propagation path cascade rate introduced in this paper is feasible for the expansion of the traditional significant profile.

The community has the effect of suppressing the spread of information. The connections between nodes with a larger out-degree are weak, while the connections between nodes with a smaller out-degree are strong, which proves that the out-degrees of nodes in the IPTs follow the power-law distribution,

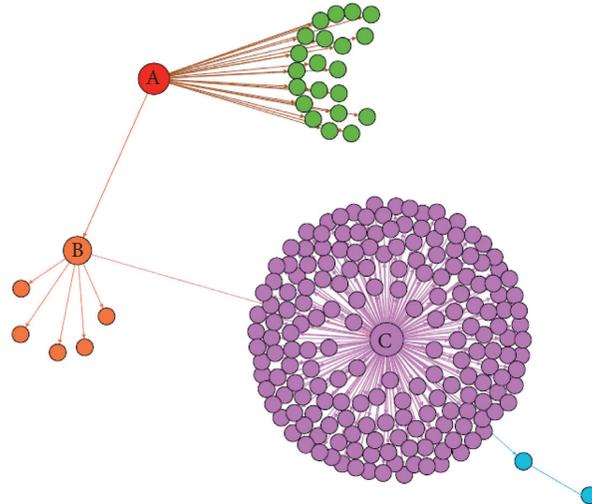


FIGURE 5: An example community structure of Sina Weibo information propagation tree. Each set of edges and nodes with same color forms a community. In fact, each community is actually a subgraph of the IPT, and it is also a tree structure. The central node of each community is often the first user to spread information in the community.

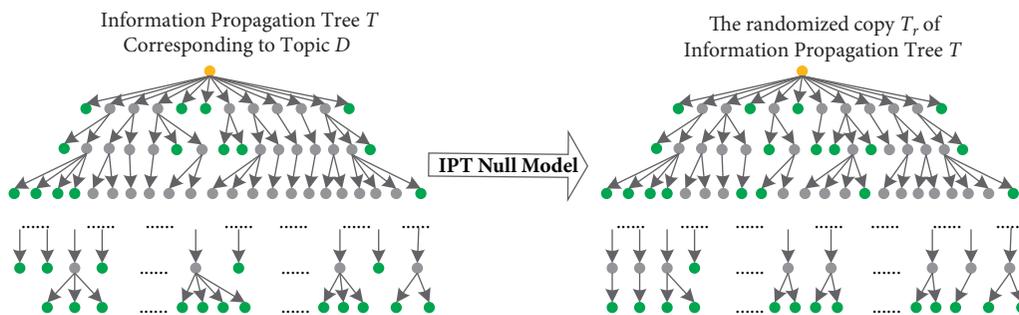


FIGURE 6: The general procedure of the IPT null model for different topics of information propagation. The left is the original IPT, and the right is the corresponding randomized copy of the IPT. The core of the edge rewiring strategy is to keep the nodes of each layer of the IPT unchanged, and then, the upper node of the two adjacent layers selects a node from the lower layer according to the probability shown in equation (3) and then establishes a directed edge.

and it is consistent with the authors' previous verified research results presented in literature [20]. This also reflects that the IPT null model can analyze the diffusion characteristics without considering the information topic attributes.

In summary, the IPT null model cannot only analyze the influence of community on information propagation but can also ignore network structures and topics to complete relevant analysis tasks, thus creating a novel direction of online social networks analysis.

6. Conclusion

In this paper, we conducted a null model empirical analysis on IPTs constructed from the Sina Weibo so as to study the effects of community structures on information propagation. We first built the null model to generate the randomized copy of Sina Weibo IPTs; then, the triad percolation-based community detection method was employed to discover the community structures in original IPTs and their corresponding randomized copy; finally, the traditional significant profile was extended in terms of R -value and Z -value to

measure the null model performance. The results of the analysis reveal that community structures constrain information propagation. In future work, we plan to devise a null model method that could extend this analysis to dynamic social networks.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded in part by the National Natural Science Foundation of China (Grant no. 61702355), Natural Science Foundation of Anhui Province (Grant no. 1908085QF283), Key Research and Development Plan of

Anhui Province (Grant nos. 202004b11020023, 202004a06020045, and 202004a05020043), Key Natural Science Project of the Anhui Provincial Education Department (Grant nos. KJ2019A0668, KJ2019A1001, and KJ2020A0733), Doctoral Research Start-Up Foundation (Grant no. 2019jb08), Overseas Visiting Project of Outstanding Young and Talents in Anhui Province (Grant no. gxgwx2020063), Domestic Visiting Project of Outstanding Young and Talents in Anhui Province (Grant no. gxgnfx2021154), Open Research Fund of National Engineering Research Center for Agro-Ecological Big Data Analysis and Application (Grant no. AE2019010), Candidates for Academic and Technical Leaders (Grant nos. 2018XJHB07 and 2019XJZY23), and Key Scientific Research Project of Suzhou University (Grant no. 2019yzd05).

References

- [1] G. Li, J. Wu, Z. Qiao, C. Zhou, H. Yang, and Y. Hu, "Collaborative social group influence for event recommendation," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM'16)*, pp. 1941–1944, Indianapolis, IN, USA, October 2016.
- [2] J. Wu, X. Zhu, C. Zhang, and Z. Cai, "Multi-instance multi-graph dual embedding learning," in *Proceedings of the 2013 IEEE 13th International Conference on Data Mining (ICDM'13)*, pp. 827–836, Dallas, TX, USA, December 2013.
- [3] J. Wu, Z. Cai, and X. Zhu, "Self-adaptive probability estimation for naive bayes classification," in *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN'13)*, pp. 1–8, Dallas, TX, USA, August 2013.
- [4] C. Liu, C. Zhou, J. Wu, Y. Hu, and L. Guo, "Social recommendation with an essential preference space," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI'18)*, pp. 346–353, New Orleans, LA, USA, February 2018.
- [5] Z. Wang, M. Jusup, H. Guo et al., "Communicating sentiment and outlook reverses inaction against collective risks," *Proceedings of the National Academy of Sciences*, vol. 117, no. 30, pp. 17650–17655, 2020.
- [6] S. Yang, J. Wang, B. Dai, X. Li, Y. Jiang, and Y. Liu, "State of the art in social network user behaviors and its future," *Bulletin of the Chinese Academy of Sciences*, vol. 30, no. 2, pp. 200–215, 2015.
- [7] Z. He, L. Wu, X. Chen, and T. Lu, "Impact of online community structure on information propagation: empirical analysis and modeling," *Journal of Harbin Institute of Technology*, vol. 3, pp. 124–128, 2013.
- [8] S. Jain, G. Mohan, and A. Sinha, "Network diffusion for information propagation in online social communities," in *Proceedings of the 2017 Tenth International Conference on Contemporary Computing (IC3)*, pp. 1–3, Noida, India, August 2017.
- [9] Y. Zhao, S. Li, and F. Jin, "Identification of influential nodes in social networks with community structure based on label propagation," *Neurocomputing*, vol. 210, pp. 34–44, 2016.
- [10] R. Zhang and D. Li, "Rumor propagation on networks with community structure," *Physica A: Statistical Mechanics and its Applications*, vol. 483, pp. 375–385, 2017.
- [11] Z. Zhang, L. Cui, Z. Pan, A. Fang, and H. Zhang, "A triad percolation method for detecting communities in social networks," *Data Science Journal*, vol. 17, no. 30, pp. 1–12, 2018.
- [12] Z. Zhang and Z. Wang, "Mining overlapping and hierarchical communities in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 421, pp. 25–33, 2015.
- [13] N. Zhao and X. Liu, "Information propagation in social networks with overlapping community structure," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 12, pp. 5927–5942, 2017.
- [14] F. Nian, L. Luo, X. Yu, and X. Guo, "Community detection in social networks based on information propagation and user engagement," *International Journal of Modern Physics B*, vol. 35, no. 8, Article ID 2150119, 2021.
- [15] J. Xu, Y. Yang, F. Jiang, and S. Jin, "Social network structure feature analysis and its modelling," *Bulletin of the Chinese Academy of Sciences*, vol. 30, no. 2, pp. 216–228, 2015.
- [16] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2658–2663, 2004.
- [17] M. Jalili and M. Perc, "Information cascades in complex networks," *Journal of Complex Networks*, vol. 5, no. 5, pp. 665–693, 2017.
- [18] L. Pellis, F. Ball, S. Bansal et al., "Eight challenges for network epidemic models," *Epidemics*, vol. 10, pp. 58–62, 2015.
- [19] 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.
- [20] Z. Zhang and Z. Wang, "The data-driven null models for information dissemination tree in social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 484, pp. 394–411, 2017.
- [21] Z. Zhang, L. Cui, A. Fang, Z. Pan, Z. Zhang, and H. Zhang, "Information dissemination analysis using a time-weight null model: a case study of sina micro-blog," *IEEE Access*, vol. 6, pp. 71181–71193, 2018.
- [22] J. Cannarella and J. A. Spechler, "Epidemiological modeling of online social network dynamics," 2014, <https://arxiv.org/abs/1401.4208v1>.
- [23] F. Chierichetti, S. Lattanzi, and A. Panconesi, "Rumor spreading in social networks," *Theoretical Computer Science*, vol. 412, no. 24, pp. 2602–2610, 2011.
- [24] P. G. Lind, L. R. d. Silva, J. S. Andrade, and H. J. Herrmann, "The spread of gossip in American schools," *Europhysics Letters*, vol. 78, no. 6, p. 68005, 2007.
- [25] H. J. Herrmann, D. C. Hong, and H. E. Stanley, "Backbone and elastic backbone of percolation clusters obtained by the new method of "burning,"" *Journal of Physics A: Mathematical and General*, vol. 17, no. 5, pp. 261–266, 1984.
- [26] F. Roshani and Y. Naimi, "Effects of degree-biased transmission rate and nonlinear infectivity on rumor spreading in complex social networks," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 85, no. 3, Article ID 036109, 2012.
- [27] S. A. Myers and J. Leskovec, "Clash of the contagions: cooperation and competition in information diffusion," in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining (ICDM'12)*, pp. 539–548, Brussels, Belgium, December 2012.
- [28] P. G. Lind, L. R. da Silva, J. S. Andrade, and H. J. Herrmann, "Spreading gossip in social networks," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 76, no. 3, Article ID 036117, 2007.
- [29] C. Yi, Y. Bao, and Y. Xue, "Mining the key predictors for event outbreaks in social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 447, pp. 247–260, 2016.

- [30] J. Fabrega and P. Paredes, "Social contagion and cascade behaviors on twitter," *Information*, vol. 4, no. 2, pp. 171–181, 2013.
- [31] T. Rodrigues, F. Benevenuto, M. Cha, K. Gummadi, and V. Almeida, "On word-of-mouth based discovery of the web," in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, pp. 381–396, Berlin, Germany, November 2011.
- [32] D. Liben-Nowell and J. Kleinberg, "Tracing information flow on a global scale using internet chain-letter data," *Proceedings of the National Academy of Sciences of United States*, vol. 105, no. 12, pp. 4633–4638, 2008.
- [33] B. Golub and M. O. Jackson, "Using selection bias to explain the observed structure of internet diffusions," *Proceedings of the National Academy of Sciences of United States*, vol. 107, no. 24, pp. 10833–10836, 2010.
- [34] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, 2002.
- [35] S. Maslov, K. Sneppen, and A. Zaliznyak, "Detection of topological patterns in complex networks: correlation profile of the internet," *Physica A: Statistical Mechanics and its Applications*, vol. 333, pp. 529–540, 2004.
- [36] D. J. D. S. Price, "Networks of scientific papers," *Science*, vol. 149, no. 3683, pp. 510–515, 1965.
- [37] D. J. D. S. Price, "A general theory of bibliometric and other cumulative advantage processes," *Journal of the Association for Information Science and Technology*, vol. 27, no. 5, pp. 292–306, 1976.
- [38] G. Palla, I. Derenyi, I. J. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.
- [39] H. Shen, X. Cheng, K. Cai, and M. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706–1712, 2009.
- [40] V. D. Blondel, J.-L. 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.