

Role of Complex Networks in Modelling Epidemic Dynamics

Lead Guest Editor: Ping Li

Guest Editors: Xiaoke Xu, Peng Ji, Ye Wu, and Spase Petkoski





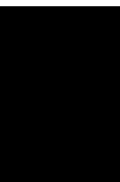
Role of Complex Networks in Modelling Epidemic Dynamics

Complexity

Role of Complex Networks in Modelling Epidemic Dynamics

Lead Guest Editor: Ping Li


Guest Editors: Xiaoke Xu, Peng Ji, Ye Wu, and Spase
Petkoski



Copyright © 2021 Hindawi Limited. All rights reserved.

This is a special issue published in "Complexity." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chief Editor

Hiroki Sayama , USA

Associate Editors

Albert Diaz-Guilera , Spain
Carlos Gershenson , Mexico
Sergio Gómez , Spain
Sing Kiong Nguang , New Zealand
Yongping Pan , Singapore
Dimitrios Stamovlasis , Greece
Christos Volos , Greece
Yong Xu , China
Xinggang Yan , United Kingdom

Academic Editors

Andrew Adamatzky, United Kingdom
Marcus Aguiar , Brazil
Tarek Ahmed-Ali, France
Maia Angelova , Australia
David Arroyo, Spain
Tomaso Aste , United Kingdom
Shonak Bansal , India
George Bassel, United Kingdom
Mohamed Boutayeb, France
Dirk Brockmann, Germany
Seth Bullock, United Kingdom
Diyi Chen , China
Alan Dorin , Australia
Guilherme Ferraz de Arruda , Italy
Harish Garg , India
Sarangapani Jagannathan , USA
Mahdi Jalili, Australia
Jeffrey H. Johnson, United Kingdom
Jurgen Kurths, Germany
C. H. Lai , Singapore
Fredrik Liljeros, Sweden
Naoki Masuda, USA
Jose F. Mendes , Portugal
Christopher P. Monterola, Philippines
Marcin Mrugalski , Poland
Vincenzo Nicosia, United Kingdom
Nicola Perra , United Kingdom
Andrea Rapisarda, Italy
Céline Rozenblat, Switzerland
M. San Miguel, Spain
Enzo Pasquale Scilingo , Italy
Ana Teixeira de Melo, Portugal



Shahadat Uddin , Australia
Jose C. Valverde , Spain
Massimiliano Zanin , Spain

Contents



Time-Varying Spatial Memory Model and Its Impact on Virus Spreading

Mingyue Xu , Dingding Han , Kaidi Zhao , and Qingqing Yao
Research Article (15 pages), Article ID 6687168, Volume 2021 (2021)

Global Dynamics of an SIS Model on Metapopulation Networks with Demographics

Maoxing Liu , Xinjie Fu , Jie Zhang , and Donghua Zhao 
Research Article (9 pages), Article ID 8884236, Volume 2021 (2021)


A Heterogeneous Branching Process with Immigration Modeling for COVID-19 Spreading in Local Communities in China

Lin Zhang , Haochen Wang, Zhongyang Liu, Xiao Fan Liu , Xin Feng , and Ye Wu
Research Article (11 pages), Article ID 6686547, Volume 2021 (2021)


Understanding the Anticontagion Process and Reopening of China during COVID-19 via Coevolution Network of Epidemic and Awareness

Lingbo Li , Ying Fan , An Zeng , and Zengru Di
Research Article (11 pages), Article ID 6623427, Volume 2021 (2021)

Recent Progress about Flight Delay under Complex Network

Tang Zhixing, Huang Shan, and Han Songchen 
Review Article (18 pages), Article ID 5513093, Volume 2021 (2021)

Assessing the Effectiveness of Mass Testing and Quarantine in the Spread of COVID-19 in Beijing and Xinjiang, 2020

Feng Li, Zhen Jin , and Juan Zhang
Research Article (10 pages), Article ID 5510428, Volume 2021 (2021)


Coupled Dynamic Model of Resource Diffusion and Epidemic Spreading in Time-Varying Multiplex Networks

Ping Huang, Xiao-Long Chen, Ming Tang, and Shi-Min Cai 
Research Article (11 pages), Article ID 6629105, Volume 2021 (2021)

A New Coupled Awareness-Epidemic Spreading Model with Neighbor Behavior on Multiplex Networks

Chao Zuo, Anjing Wang, Fenping Zhu, Zeyang Meng, and Xueke Zhao 
Research Article (14 pages), Article ID 6680135, Volume 2021 (2021)



Spreading Dynamics of a 2SIH2R, Rumor Spreading Model in the Homogeneous Network

Yan Wang, Feng Qing , Jian-Ping Chai, and Ye-Peng Ni
Research Article (9 pages), Article ID 6693334, Volume 2021 (2021)

Evolving Nature of Human Contact Networks with Its Impact on Epidemic Processes

Cong Li , Jing Li , and Xiang Li 
Research Article (13 pages), Article ID 6643658, Volume 2021 (2021)

Policy and Law Assessment of COVID-19 Based on Smooth Transition Autoregressive Model

Jieqi Lei, Xuyuan Wang, Yiming Zhang, Lian Zhu , and Lin Zhang 




Research Article (13 pages), Article ID 6659117, Volume 2021 (2021)

The Artificial University: Decision Support for Universities in the COVID-19 Era

Wesley J. Wildman , Saikou Y. Diallo , George Hodulik , Andrew Page , Andreas Tolk , and Neha Gondal 

Research Article (10 pages), Article ID 5910209, Volume 2020 (2020)

Risk Assessment of Biological Asset Mortgage Loans of China's New Agricultural Business Entities

Shuzhen Zhu , Yutao Chen , and Wenwen Wang 

Research Article (12 pages), Article ID 8865840, Volume 2020 (2020)

Research Article

Time-Varying Spatial Memory Model and Its Impact on Virus Spreading

Mingyue Xu ¹, Dingding Han ¹, Kaidi Zhao ¹ and Qingqing Yao²

¹*School of Information Science and Technology, Fudan University, Shanghai, China*

²*School of Communication and Electronic Engineering, East China Normal University, Shanghai, China*

Correspondence should be addressed to Dingding Han; ddhan@fudan.edu.cn

Received 16 October 2020; Revised 24 February 2021; Accepted 13 September 2021; Published 4 October 2021

Academic Editor: Ping Li

Copyright © 2021 Mingyue Xu 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 models of time-varying network have a profound impact on the study of virus spreading on the networks. On the basis of an activity-driven memory evolution model, a time-varying spatial memory model (TSM) is proposed. In the TSM model, the cumulative number of connections between nodes is recorded, and the spatiality of nodes is considered at the same time. Therefore, the active nodes tend to connect the nodes with high intimacy and close proximity. Then, the TSM model is applied to epidemic spreading, and the epidemic spreading on different models is compared. To verify the universality of the TSM model, this model is also applied to rumor spreading, and it is proved that it can also play a good inhibiting effect. We find that, in the TSM network, the introduction of spatiality and memory can slow down the propagation speed and narrow the propagation scope of disease or rumor, and memory is more important. We then explore the impact of different prevention and control methods on pandemic spreading to provide reference for COVID-19 management control and find when the activity of node is restricted, the spreading will be controlled. As floating population has been acknowledged as a key parameter that affects the situation of COVID-19 after work resumption, the factor of population mobility is introduced to calculate the interregional population interaction rate, and the time-varying interregional epidemic model is established. Finally, our results of infectious disease parameters based on daily cases are in good agreement with the real data, and the effectiveness of different control measures is evaluated.

1. Introduction

The transmission period of COVID-19 coincided with the Spring Festival travel rush on the eve of the Chinese New Year, during which a large population movement occurred throughout the country, expanding the spread of the virus and accelerating its spread [1, 2]. The Chinese government adopted series of control measures to cut off the transmission path of the virus. After February 4, the number of reported daily cases in China began to decrease, indicating that the measures were working. Because of the certain negative impact on the lives of residents and social economy, some regions and enterprises began to return to work after February 10, and the population mobility rate rose. The flowing of population made it more difficult to control the epidemic, so it was of great significance to study the way and modes of transmission of COVID-19.

From previous studies, we can learn about some epidemiological and clinical characteristics of COVID-19 [3, 4]. Compared with SARS and H1N1, the transmission of COVID-19 is stronger. The infection source is mainly a patient infected by COVID-19. A number of studies have proved that asymptomatic infected persons may also be the source of infection and that person-to-person contact is the main cause of virus transmission. Through analysis of early cases of COVID-19, Li et al. found that only 8.6% of the cases developed after January 1 were related to the seafood wholesale markets in South China, and human-to-human transmission had occurred among close contacts since mid-December 2019 [5]. Previous research has confirmed that contact is the main way for the spread of the virus, but in order to prevent and control the transmission of COVID-19 more effectively, further modeling and research are needed for the transmission dynamics of the virus.

Many scholars have studied the development and policies of COVID-19 for further prevention and control from the perspective of epidemic dynamics [6, 7]. Jia et al. [8] used the location data of mobile phone to track the flow of people from Wuhan to other areas and analyzed the transmission process of the pandemic based on the data. Kraemer et al. [9] used mobility data from Wuhan and case data including travel history to analyze the influence of travel restrictions on the spread of COVID-19 in China and found that travel restrictions in the early stages of the pandemic could be effective in preventing the importation of the virus. Moreover, complex network is an important tool for simulating transmission dynamics of virus or information on networks and can provide theoretical basis and guidance for it. COVID-19 has been analyzed using complex networks by some scholars [10]. Basnarkov [11] studied the Susceptible-Exposed-Asymptomatic-Infectious-Recovered (SEAIR) epidemic spreading model to capture two characteristics: delayed start and its appearance before onset of symptoms. Some scholars have used multilayer networks to analyze the COVID-19 epidemic from the individual characteristics of different countries, such as social activities and economic characteristics [12, 13]. Wang et al. [14] proposed a disease model based on two-layered multiplex networks, demonstrating that the epidemic prevalence and the epidemic outbreaks can be suppressed by the diffusion of positive preventive information and be promoted by the diffusion of negative preventive information. In addition, the dynamics of multiple information transmission and epidemic co-evolution under the influence of mass media are discussed in [15].

The spread of COVID-19 is mostly analyzed from the perspective of overall population flow for epidemic prevention and control measures, but the establishment and transmission mode of interpersonal social network of COVID-19 spreading are lacking. In this paper, from the construction of a disease transmission network, we try to study COVID-19 from the perspective of transmission dynamics on the time-varying network.

In most real networks, individual attributes, such as activity and location, all change over time and lead to changes in network topologies on the macroscopic view. Therefore, it is more appropriate to use time-varying networks to simulate the structure of real networks [16]. Time-varying networks are widely used [17–19]. According to characteristics of social networks, N perra et al. verified degree distribution and weight distribution of nodes in the aggregated static network met the power-law distribution. The ratio of the number of activities of node i to the number of activities of all nodes in a period of time was defined as the activity of node i , and it was found that the distribution of activity also met the power-law distribution. Therefore, N perra et al. proposed the activity-driven model to describe time-varying networks [20]. Many scholars have proposed improvements and optimizations to the time-varying network model, to make it more in line with the real network and dynamic characteristics, and mainly applied it to the epidemic transmission and the information transmission in social networks [21]. Some scholars set properties for nodes,

such as attractiveness and willingness [22–29] or introduced external factor to the process of propagation [30–32].

We first try to build a network model closer to the real network. In the real network, the process of node connecting is not completely random, but the result of the non-Markov process with memory. Especially in social networks, social reinforcement derived from memory is the main feature of social contagions. It is of great practical significance to add memory mechanism [33–40] in time-varying networks. Moreover, it is known that, in our daily life, we tend to have more contact with the people around us or those close to us, while there is little work to consider spatial factors on the basis of the activity model.

Considering these characteristics, spatiality, activity, and memory of node are introduced into the construction of network and a time-varying spatial memory (TSM) model is proposed, which can better reflect real network characteristics. The TSM model is applied to the epidemic spreading to provide a reference for the spread, management, and control of the COVID-19 epidemic. Considering that infectious disease is the virus in biology, while rumor is the virus in sociology, this model is also applied to rumor spreading to prove the universality of the TSM model. Moreover, a time-varying interregional epidemic model under the influence of floating population is proposed to explore the impact of interregional population after returning to cities, and the influence and effect of different epidemic control measures are compared.

The remaining sections in this paper are organized on the basis of the following organization. In Section 2, a time-varying spatial memory model considering spatiality and memory of nodes is proposed based on the RP model. In Section 3, SEIR spreading and its simulation of the propagation of COVID-19 under the influence of floating population are studied. The TSM model is applied to the spread of the epidemic and also rumor, and the effect of control measures is analyzed in Section 4. In Section 5, an empirical study on the impact of returning to work in the epidemic situation is conducted. Last, conclusion is explained.

2. Time-Varying Spatial Memory Network Model Based on the RP Model

A model based on activity driven is proposed by N Perra et al., introducing the memory of nodes [40]. In this model, each node records all nodes that have been connected with it before time t to form set R . When active nodes establish edges at time t , the probability of selecting nodes in set R increases with the increase of the number of nodes in it. However, for nodes belonging to set R , the probability of being selected is the same, which is also the limitation of the memory model. In the actual network, if the cumulative number of edges between pairs of nodes is taken as the edge weight, the weight will reflect the intimacy between nodes and have a great influence on the preference probability of connecting edges. For example, in social networks, the more the connections between individuals, the closer the relationship. Although some nodes have been connected, they

may be only accidental. Obviously, we are more likely to spread information to friends and acquaintances and we cannot simply assume that the two categories are the same as in the RP model.

Meanwhile, both the spatiality and memory of nodes in time-varying networks will affect network structure. In social networks, for example, when we want to connect with others, we tend to choose the ones closer to us. Although the online social network is currently booming, if the distance between nodes is similar, the two are often in the same community. Distance is still one of the driving factors for node connection. On the other hand, we prefer to share with our close friends and partners when we get the news, rather than randomly choose one of the people we know.

Therefore, we consider both spatiality and memory of nodes in the model of time-varying networks. Nodes do not only record whether they have established a connection with other nodes but also record the total number of interactions with other nodes from the first time step and are more inclined to connect to nodes that are closer to them. These preferences are reflected in the preference probability of the destination node selected. The network model, named as the time-varying spatial memory (TSM) model, is constructed as follows:

- (1) An activity a_i is assigned to each node at the beginning of constructing the network, which is subject to the power-law distribution.
- (2) In each discrete time step Δt , the initial state of network G_t is composed of N disconnected nodes.
- (3) Each node becomes active with probability $a_i \Delta t$. The active node creates m edges and connects to other nodes. Inactive nodes can only receive edges. The probability of node i choosing node j to establish edges is as follows:

$$P_{ij} = \begin{cases} \frac{n}{n+C} \cdot \left(\frac{W_{ij}^\alpha / d_{ij}^2}{\sum_{j \in R} (W_{ij}^\alpha / d_{ij}^2)} \right), & j \in R, \\ \frac{C}{n+C} \cdot \left(\frac{1/d_{ij}^2}{\sum_{j \notin R, j \neq i} (1/d_{ij}^2)} \right), & j \notin R, \end{cases} \quad (1)$$

where R is the set of nodes that have been connected to node i . n is the number of nodes in set R . C is the bias constant, and $C = 1$ is set here. W_{ij} is the cumulative number of edges between node pairs, d_{ij} is the Manhattan distance between node i and node j , and α is the proportion parameter. The larger the value of α is, the greater the influence of node memory on preference probability is.

As the number of connected nodes increases, the probability of connecting new nodes decreases. In [40], the author used $(C/(n+C))$ to define the probability of connecting a new node and showed that there would be a probability of $(n/(n+C))$ to connect an old node. In addition to the difference between the old and the new nodes, the probability of

connecting the given two nodes increases as their intimacy increases, and whether it is a new node or an old node, the probability of creating an edge between two nodes decreases as the distance between them increases. So, after choosing the new node or the old node by probabilities $(C/(n+C))$ and $(n/(n+C))$, we set the probability of choosing a particular node as proportional to intimacy and inversely proportional to distance. Also, we use a parameter α to adjust the ratio between intimacy and distance.

According to the formula for the probability of connecting edges,

- (1) The active node i first decides whether to select the destination node in set R according to the probability $P_{\text{oldneighbor}}(n) = (n/(n+C))$.
- (2) If so, the destination node is selected according to the intimacy and distance between each node in set R and node i . Otherwise, the closer node is randomly selected as the destination in the remaining new nodes.
- (4) At the next time step $t + \Delta t$, all edges in the network G_t are deleted and the second step is repeated to get the network $G_{t+\Delta t}$ at the next time step.

We analyze the degree distribution of the network generated by the TSM model and set $N = 1000$, $L = 32$, $m = 3$, $a < 0.8$, and $\alpha = 0.5$. A network of 2000 time steps is generated, and every 400 time steps, the degree distribution of the current time step and the degree distribution of the integrated network superimposed at all times (including and excluding repeated edges) are recorded, as shown in Figure 1. Figure 1(a) shows the degree distribution of a single time step. As can be seen from the figure, the degree distribution of each time step is similar, and nodes with a degree value of 0 account for about 50% of the total nodes. Figure 1(c) shows the degree distribution of the integrated network. As time goes by, the number of edges in the network increases, and there are fewer nodes with small or large degree value, while there are more nodes with intermediate degree value. As can be seen from the figure, because the node activity in the network is the power-law distribution, it is difficult to be active for some nodes, and their cumulative degree values change slightly over time, while the degree values of nodes with higher activity increase significantly over time. Also, because the node considers the memory attribute when selecting the destination node, the difference between different time steps becomes smaller as time goes by. Figure 1(e) shows the degree distribution including the number of repeated edges, as there are repeated edges between a pair of nodes in the integrated network. In addition, we record the distribution of degree when the activity of nodes is set as 0.1, as shown in Figures 1(b), 1(d), and 1(f). The node distribution of a single time step in Figure 1(b) is similar to that in Figure 1(a), while in Figures 1(d) and 1(f), due to the same node activity, the degree value increases evenly.

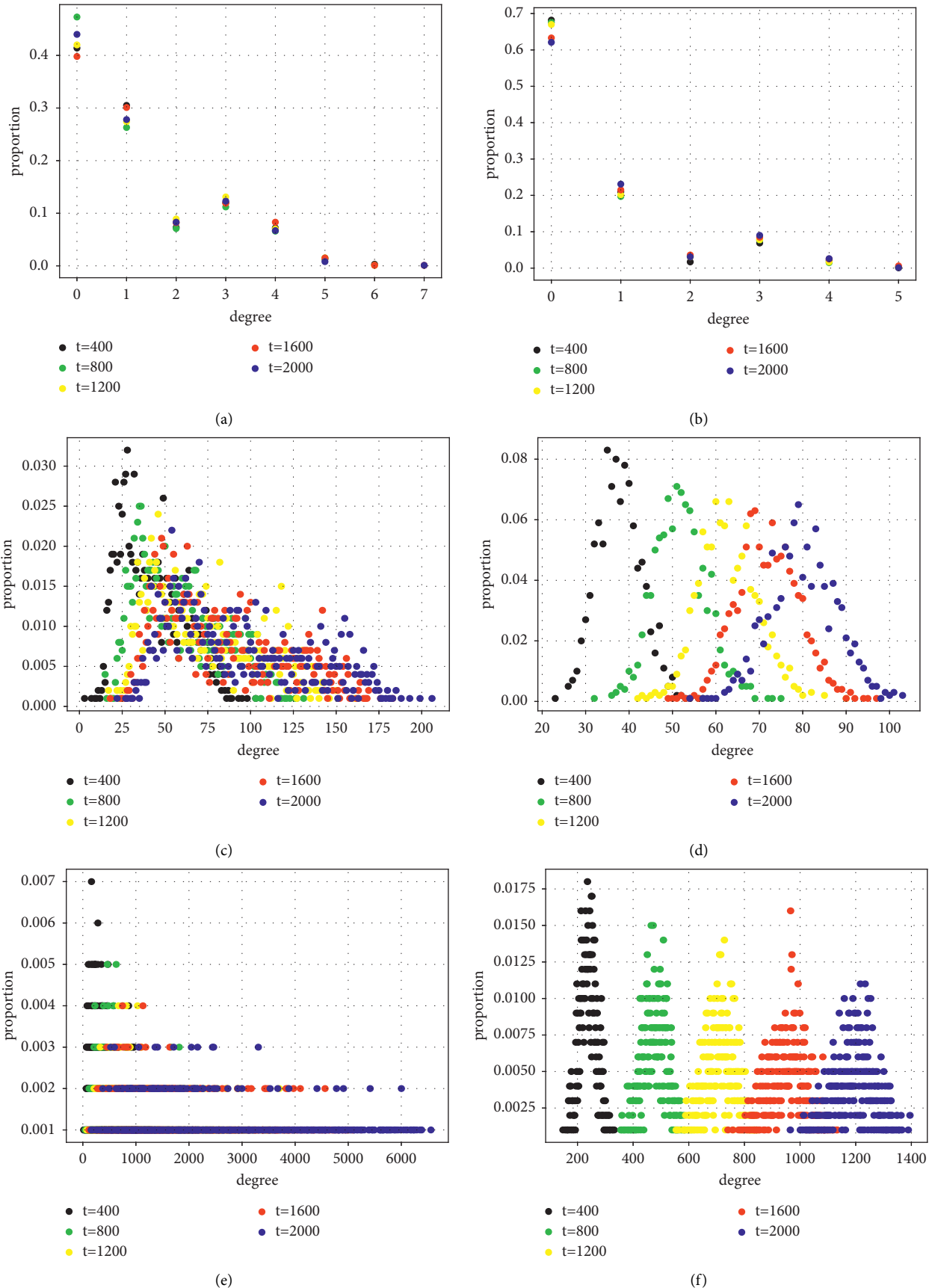


FIGURE 1: Degree distribution of the network. (a) The degree distribution of a single time step. (c) The degree distribution of the integrated network excluding repeated edges. (e) The degree distribution of the integrated network including repeated edges. $N = 1000$, $L = 32$, $m = 3$, $a < 0.8$, and $\alpha = 0.5$. (b) The degree distribution of a single time step when $a = 0.1$. (d) The degree distribution of the integrated network excluding repeated edges when $a = 0.1$. (f) The degree distribution of the integrated network including repeated edges when $a = 0.1$.

Because the TSM model comprehensively considers three driving attributes of nodes that affect the network topology and the proportion of driving factors that can be adjusted by parameter, it can be used as a general model framework for studying a complex system and its dynamic behaviors. The dynamics on the model also can apply to the spread of an epidemic and rumor.

3. SEIR Spreading

The epidemic model SEIR divides population into four categories: susceptible (S), exposed (E), infectious (I), and recover (R), and the four groups of individuals will transform or die at a certain rate. With the development of the epidemic and the improved control of COVID-19 after February 1, a variant of the SEIR model containing probable and isolated individuals [41] is adopted in this paper. $S_k, E_k, I_k, P_k, J_k, R_k$, respectively, label and quantify susceptible, exposed, infected, probable, isolated, and recovered individuals at time t . Then, the total population in the area k at time t is

$$P_{\text{total}} = S_k + E_k + I_k + P_k + J_k + R_k. \quad (2)$$

Probable individuals P_k are quarantined individuals with confirmed epidemiological contact with a source of COVID-19. It is assumed that all quarantined individuals include susceptible individuals and exposed individuals. When the exposed individuals develop symptoms, they become isolated individuals J_k , but their contact is reduced due to quarantine. So, even if they are exposed, they are not infectious, and after quarantine observation, asymptomatic groups will become susceptible groups and exposed groups will become isolated. Isolated population J_k : infected persons become the isolated population through isolation or hospitalization, losing the ability to infect the susceptible population, which will be diminished due to recovery and disease-induced death.

We also consider the influence of population mobility. We suppose there are several geographical regions in addition to the current region, each region can be regarded as a community, and people from the same community are homogeneous; that is, they have the same biological and epidemiological parameters. According to the control measures for COVID-19, the time is divided into the period of strict control from February 1 to February 10 and the period of returning to work after February 10. Suppose that population flow between different regions is 0 from February 1 to February 10 because of strict controlling. After February 10, due to the impact of work resumption, it is assumed that except the probable and the isolated who cannot leave the area, the remaining individuals can enter and leave the areas at a certain rate, but the flow rate of infected individuals is affected by epidemic control. Supposing that, after returning to work on February 10, the population inflow rate in the area k is φ_k , the population outflow rate is ρ_k and the limit coefficient of epidemic control on infected people is c . $c = 0$ indicates that infected individuals cannot leave their areas, and $c = 1$ indicates that infected individuals can move at the

same rate as the susceptible. S_l, E_l, I_l, R_l represent the total number of the susceptible, exposed, infected, and recovered individuals in other areas except region k , respectively.

The spread coefficients of exposed and infected persons are, respectively, β_E and β_I , which represent the transmission probability of COVID-19. The infection rate of susceptible people in region k is $((\beta_E E_k + \beta_I I_k)S_k/P_{\text{total}})$.

Since it is not possible to accurately distinguish COVID-19 patients and ordinary influenza patients, ordinary influenza patients turn into probable individuals by probability θ_S and recover to susceptible individuals by probability θ_P after a period of isolation. γ_E indicates the probability that an exposed individual will turn into a probable individual, and then, ε_E means the exposed individuals will turn to infected with probability ε_E . ξ_I refers to the self-healing rate of the infected individual, γ_I represents the isolation rate of the infected individual in hospital, σ represents the mortality rate of infected individuals, ε_P represents the probability of the probable individuals becoming infected and being isolated, and ξ_J indicates the cure rate of the hospitalized isolated population. We can get the equations of different groups of people in region k :

$$\begin{aligned} \frac{dS_k}{dt} &= \varphi_k S_l - \rho_k S_k - \frac{(\beta_E E_k + \beta_I I_k)S_k}{P_{\text{total}}} + \theta_P P_k - \theta_S S_k, \\ \frac{dE_k}{dt} &= \varphi_k E_l - \rho_k E_k + \frac{(\beta_E E_k + \beta_I I_k)S_k}{P_{\text{total}}} - \varepsilon_E E_k - \gamma_E E_k, \\ \frac{dI_k}{dt} &= c\varphi_k I_l - c\rho_k I_k + \varepsilon_E E_k - \gamma_I I_k - \xi_I I_k - \sigma I_k, \\ \frac{dP_k}{dt} &= \gamma_E E_k - \theta_P P_k + \theta_S S_k - \varepsilon_P P_k, \\ \frac{dJ_k}{dt} &= \varepsilon_P P_k + \gamma_I I_k - \xi_J J_k - \sigma J_k, \\ \frac{dR_k}{dt} &= \varphi_k R_l - \rho_k R_k + \xi_I I_k + \xi_J J_k. \end{aligned} \quad (3)$$

4. The Impact of Time-Varying Network Models on Transmission Dynamics

In real life, the network of contact between people changes over time. In order to explore the spread of COVID-19 on networks, the epidemic model in Section 3 is applied to the TSM model, and the influence of different time-varying network models is analyzed to prove the inhibition of the TSM model on propagation. Prevention and control strategies of epidemic on the spread of the epidemic are also analyzed.

When the epidemic model is applied to time-varying network models, the contact network between susceptible $S(t)$, exposed $E(t)$, infected $I(t)$, and recovered $R(t)$ is simulated. As we assume that the other two probable groups $P(t)$ and isolated groups $J(t)$ have little influence in the contact network, they are not discussed here.

In a time-varying network, when a susceptible (S) individual and an infected (I) one meet, the susceptible may become exposed (E) with probability β_I . If an exposed individual contacts with a susceptible one, the susceptible individual will become an exposed one with probability β_E ; each time, exposed individuals turn infectious with probability ε_E and the infectious individuals become recovered individuals with probability ξ_I . Figure 2 shows the propagation illustration of the model of epidemic spreading. In the figure, 25 nodes are taken as examples. Inactive nodes are represented by circles and active nodes are represented by stars.

The number of network nodes $N = 10^4$, $L = 100$, which means we distribute the nodes in a 100×100 grid in the simulation, the active nodes create m edges each time where $m = 3$, the minimum value of activity a is set as 0.1, the maximum value as 1.0, and the distribution of activity follows the power-law distribution with a power exponent of 0.78. Without loss of generality, we set $\beta_I = 0.80$, $\beta_E = 0.08$, $\varepsilon_E = 0.10$, $\xi_I = 0.10$, and $\alpha = 0.50$ to get a clear view of the epidemic spreading.

The spread of the epidemic on the TSM model is simulated, as shown in Figure 3. There is a peak in the proportion of exposed and infected individuals. After 100 time steps, the proportion of the susceptible individuals decreases from 1 to 0 and the proportion of the recovered individuals increases from 0 to 1.

To compare the difference of epidemic spreading between the TSM model and other time-varying network models and prove the inhibition of the TSM model on virus propagation, the process of epidemic spreading is also carried out on the ML model, IRP model considering only memory, and AD model considering only distance, respectively.

The probability of node i selecting node j of the IRP model is as follows:

$$P_{ij} = \begin{cases} \frac{n}{n+C} \cdot \frac{W_{ij}}{\sum_{j \in R} W_{ij}}, & j \in R, \\ \frac{C}{n+C} \cdot \frac{1}{b}, & j \notin R, \end{cases} \quad (4)$$

where b is the number of nodes that node i has not connected.

The probability of node i selecting node j of the AD model is as follows:

$$P_{ij} = \frac{1}{d_{ij}^2}. \quad (5)$$

The proportion of exposed people $E(t)$ in networks is shown in Figure 4. The AD model only considers the distance between nodes and only reduces the spread of the disease to a little extent. The IRP model considers the influence of memory and records the number of connections as the intimacy between nodes, which delays the spread of the epidemic and reduces the range of spreading. The TSM model reflects the contact feature more accurately because people tend to move around or get in touch with their close friends in general. Among the time-varying networks generated by the three models, the TSM network has the slowest

spread of the epidemic and effectively suppresses it. In Figure 4, compared with the ML network reaching 57.14% of the maximum proportion of exposed persons, the TSM network reaches the maximum spread of 44.03%. The proportion of exposed persons is reduced by about 13%. This is because when a node chooses the destination node, it first divides the nodes into two categories by memory, which improves the probability of choosing old neighbors. At this time, the intimacy and distance between nodes work together, which makes the preference of connect stronger and the randomness weaker, and the ‘‘strong connection’’ and ‘‘weak connection’’ in the network become more obvious.

After the outbreak of COVID-19, a series of management and control measures were taken in China. One of the most important measures was to reduce the frequency of residents going out, to avoid the movement of people between regions, and to reduce personal contacts. In order to explore the effect of reducing human contact, different active mechanisms of nodes are simulated. First, the activity of nodes in the network is reduced. The maximum value is 0.8, and the distribution of activity still follows a power-law distribution with a power exponent of 0.78, making it more difficult for nodes to be active and reducing the number of active nodes to a certain extent. Figure 5(a) shows the proportion of infected nodes in the network. When the node activity is reduced and it is no longer easy to contact people, the proportion of infected people in the network is significantly reduced, and the time of the outbreak is delayed, from the maximum value of 31.7% to 21.61%. The epidemic has been effectively prevented and controlled, which reflects the inhibition of strong connection.

Then, another possible situation about node activity is considered. The minimum value of activity a is set to 0.1, and the distribution of activity still follows a power-law distribution with a power exponent of 0.78, while the number of edges generated by active nodes is set to $m = 1, 2, 3, 4, 5, 6$, respectively. The results are shown in Figure 5(b). When the number of edges created by active nodes is reduced, fewer people are contacted in each time step, which also slows down the spread of the epidemic and the scope becomes smaller. When m is set to 1, the range of spread is reduced to about 15%, and the peak of the outbreak is greatly delayed.

Due to rapid advances in medical technology for COVID-19, the cure rate of COVID-19 patients within a certain period of time has been effectively improved. Therefore, we assume that the probability of an infected person becoming recovered person increases to 0.3 at each time step, and the rate of infected people is compared with that of $\xi_I = 0.1$. The results are shown in Figure 5(c). Although the outbreak of the epidemic is hardly delayed, the number of people recovering from the infected increases at every moment, which has greatly reduced the number of infected people. Therefore, improving the cure probability of infected persons is also crucial to control the epidemic.

In order to explore the influence of the ratio of spatiality and memory in the TSM model on epidemic transmission, we compare the ratio of susceptible individuals in epidemic spreading under different α , and the results are shown in Figure 6. It can be seen that the greater the value of α (i.e., the

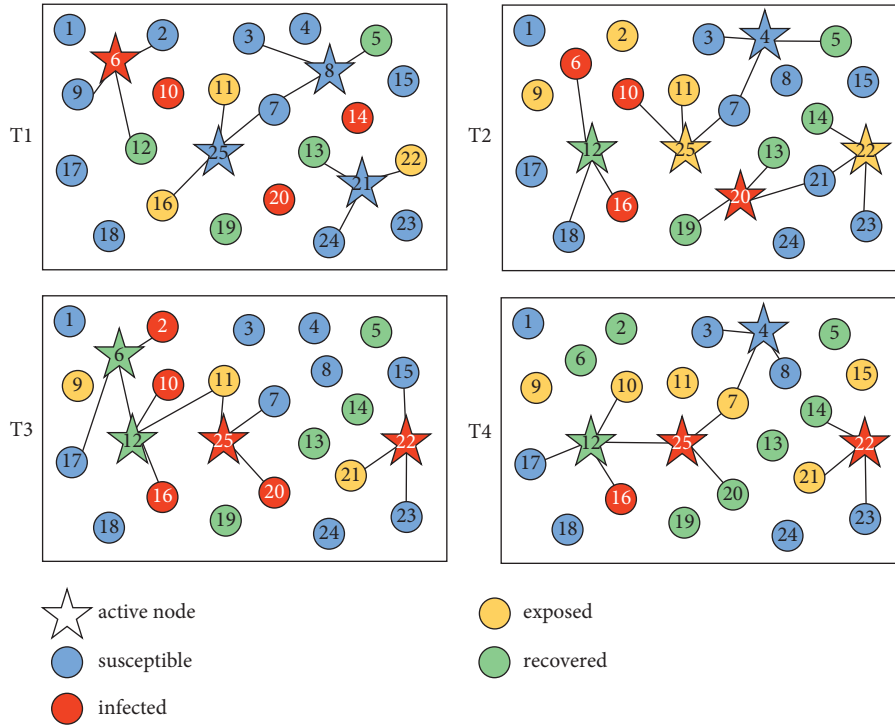


FIGURE 2: Schematic illustration of epidemic spreading on the TSM model. When a susceptible (S) individual and an infected (I) one meet, the susceptible may become exposed (E) with probability β_I . If an exposed individual contacts with a susceptible one, the susceptible individual will become an exposed one with probability β_E ; each time, exposed individuals turn infectious with probability ε_E and the infectious individuals become recovered individuals with probability ξ_I . T1 means the first time step, T2 means the second time step, and so on. 25 nodes are taken as examples in the illustration. The distribution of activity follows power-law distribution, as some nodes are often active, while others are rarely active.

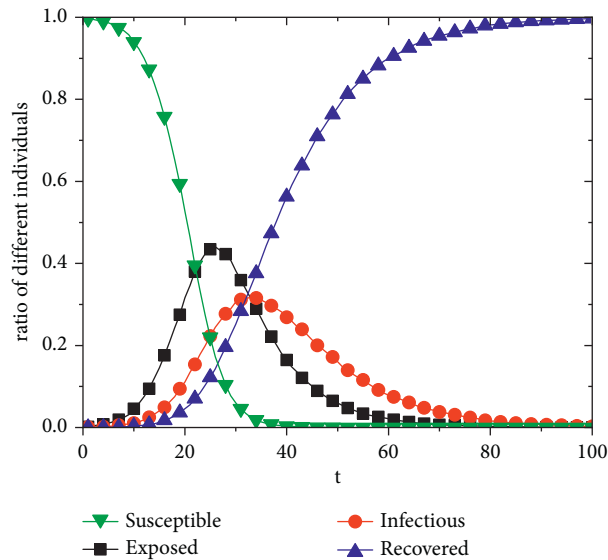


FIGURE 3: Behaviors of the epidemic spreading on the TSM model. There is a peak in the proportion of exposed and infected individuals. The proportion of susceptible individuals decreases from 1 to 0, and the proportion of recovered individuals increases from 0 to 1. $N = 10^4$, $L = 100$, $m = 3$, $0.1 < a < 1.0$, $\beta_I = 0.80$, $\beta_E = 0.08$, $\varepsilon_E = 0.10$, $\xi_I = 0.10$, and $\alpha = 0.50$, and the distribution of activity follows a power-law distribution with a power exponent of 0.78.

greater the dominance of memory over connection in the TSM model), the stronger the speed, and the scope of the epidemic spreading is inhibited. Specifically, when $\alpha = 0$, the

node selects the destination node in set R only considering the distance. The range of epidemic spreading is close to 96%. When $\alpha > 0$, the intimacy affects the selection of the

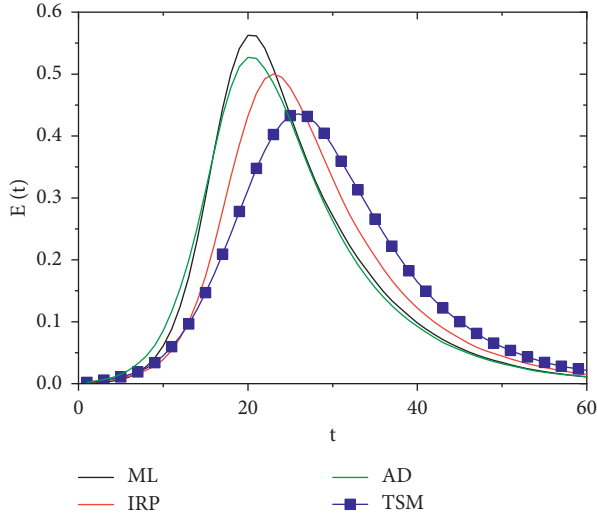


FIGURE 4: Comparison of proportions of exposed persons on time-varying networks generated by the four models. $N = 10^4$, $L = 100$, $m = 3$, $0.1 < a < 1.0$, $\beta_I = 0.80$, $\beta_E = 0.08$, $\varepsilon_E = 0.10$, $\xi_I = 0.10$, and $\alpha = 0.50$, and the distribution of activity follows a power-law distribution with a power exponent of 0.78.

edge. With the increase of α , the proportion of intimacy increases, and the inhibition effect on the speed and range of epidemic spreading is also gradually enhanced. For example, when $\alpha = 4$, the spread range drops to about 66%. Thus, it can be further verified that memory has a greater inhibitory effect on transmission compared with spatiality.

In order to verify whether the epidemic spreading on the time-varying network is consistent with the real epidemic spreading of COVID-19, the parameters are adjusted and the number of infected persons is compared with the real number of confirmed cases in Shanghai, Zhejiang, and Henan from February 1 to 29 [42]. It can be seen from Figure 7 that the propagation trend in time-varying networks and real networks is roughly the same. Because we adjust the parameters of different regions in the model, the transmission process is more in line with the real network. The number of infected people goes through a peak from low to high, and then, it declines with most of the people becoming immune. It is of great practical significance to study the propagation process on time-varying networks.

In order to further prove the universality of the TSM model, we also apply the model to the rumor propagation process in social networks. The spread of rumors is similar to the spread of viruses in some ways. Infectious disease is the virus in biology, while rumor is the virus in sociology.

There are three states of nodes in the network, namely, ignorant, spreader, and recover. In each time step, if the spreader interacts with the ignorant, the ignorant will become a spreader with the infection rate λ . In the event of an interaction between two spreaders or an interaction between the spreader and the recover, the person may be aware of the rumor or find that the information is known to the public, so the interest in continuing to transmit the information will decline and the spreader will become a recover at the recovery rate μ . We set the parameters of the models as

$N = 10^5$, $m = 1$, and the distribution of activity follows a power-law distribution with a power exponent of 2.8, $\lambda = 1$, and recovery rate $\mu = 0.6$. The rumor spreading process is observed on the network generated by the TSM model, and the results are compared with the ML model, AD model, and IRP model. Figure 8 shows the proportion of ignorant $I_g(t)$ with time step. The influence of each driving factor on propagation speed and range is analyzed.

The results of rumor propagation on networks are similar with that of the epidemic spreading. The results based on AD and ML are very similar. The rumor propagation speed in the AD model drops slightly, and there is basically no impact on the spreading ranges. The IRP model has an obvious constraint on the rumor spreading. In the TSM model, the propagation speed and range are further controlled. We adjust the parameters of the models to simulate a real mail network. The mail network is built from the mail exchange data of about 1000 staff in a European research institute in 803 days, and the data are from Stanford University's data website SNAP. The rumor propagation on various networks is shown in the subgraph of Figure 8. The network generated by the TSM model is the closest to the result of the actual network among several models.

Figure 9 shows the rumor propagation on the network generated by the TSM model when the proportion parameter α takes different values with the number of nodes $N = 10^4$. Different with epidemic spreading, multiple connections between two nodes may cause the node to become recover more quickly. It may lead to the stronger inhibition effect of memory in the results.

5. Impact of Work Resumption on COVID-19

We collect the number of daily confirmed, existing confirmed, suspected, cured, and dead cases of COVID-19 in China as of February, 2020. The data come from reported cases in Chinese provinces online [42]. We estimate the parameters in the model by existing data and referring to existing work.

First, the parameters of the model from February 1 to 10 are estimated according to the work in [5, 43, 44] and the data, and the following parameters are used for simulation: $\beta_I = 0.1379$, $\beta_E = 0.01379$, $\theta_P = 0.04002$, $\theta_S = 0.02213$, $\varepsilon_E = 0.02083$, $\varepsilon_P = 0.2$, $\gamma_E = 0.1875$, $\gamma_I = 0.3448$, $\xi_I = 0.002$, $\xi_J = 0.03$, $\sigma = 0.0021$. We assume the epidemic control is strict during the period from February 1 to February 10; then, $\varphi_k = 0$, $\rho_k = 0$, $c = 0$. Numerical simulation of daily COVID-19 patients in Shanghai is shown in Figure 10.

Although strong prevention and control measures have played an important role in curbing the outbreak of COVID-19, these measures have also had a huge negative impact on daily life and economic development. Many regions and enterprises start work resumption after February 10, which also means that the population flow begins to increase and may lead to a renewed COVID-19 outbreak. To solve this problem, a simulation is carried out by controlling the daily inflow and outflow of population. δ represents the ratio of daily inflow and outflow to the local population. Assuming

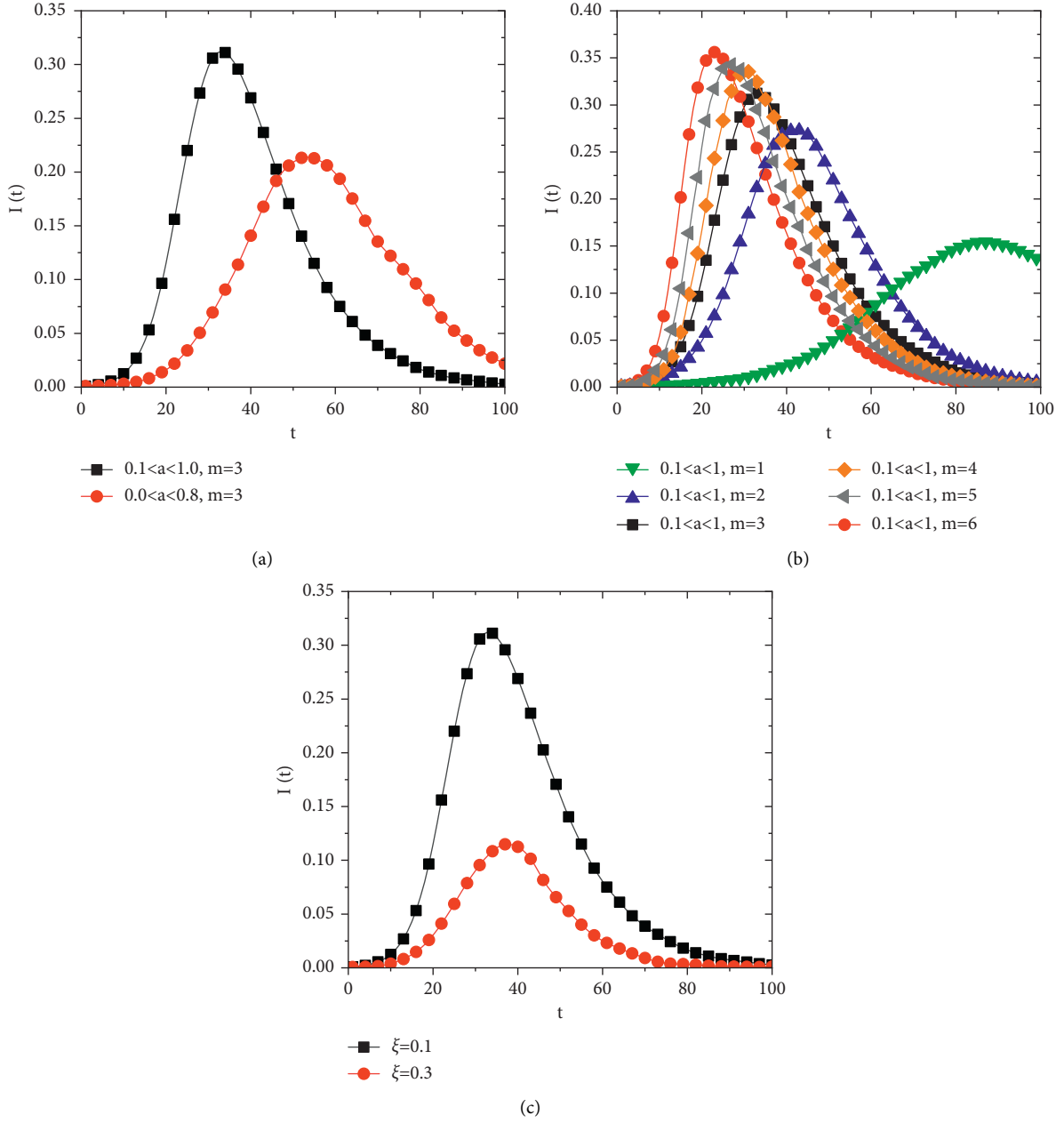


FIGURE 5: Comparison of infected persons under different management and control measures. (a) Proportion of infected persons before and after reducing node activity. The activity a is set to $0.1 < a < 1.0$ and $0.0 < a < 0.8$, respectively, and the distribution of activity still follows a power-law distribution with a power exponent of 0.78. (b) Proportion of infected persons with different number of edges created by active nodes, $m = 1, 2, 3, 4, 5, 6$, respectively. (c) Proportion of infected persons before and after improving the recovery probability, $\xi_I = 0.1, 0.3$, respectively.

that the distribution of various groups of people in various regions is similar, there are

$$\delta = \frac{\varphi_k S_l - \rho_k S_k}{S_k} = \frac{\varphi_k E_l - \rho_k E_k}{E_k} = \frac{\varphi_k R_l - \rho_k R_k}{R_k}. \quad (6)$$

The movement of infected people δ_I is affected by epidemic control:

$$\delta_I = \frac{c\varphi_k I_l - c\rho_k I_k}{I_k} = c\delta, \quad (7)$$

where $S_k, E_k, I_k, P_k, J_k, R_k$, respectively, represent susceptible, exposed, infected, probable, isolated, and recovered individuals in the area k , S_l, E_l, I_l, R_l represent the total number of susceptible, exposed, infected, and recovered individuals in other areas except region k , respectively. The population

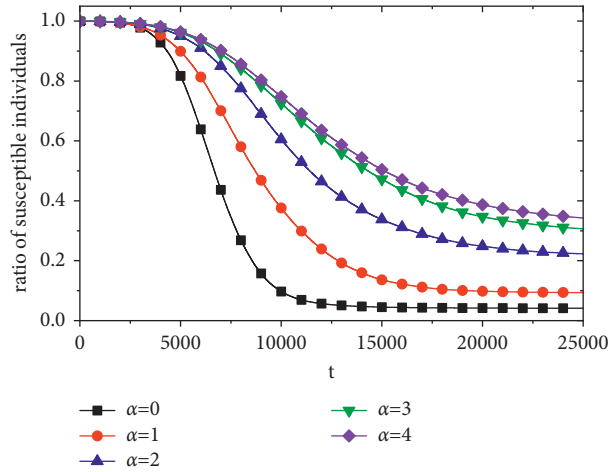


FIGURE 6: The impact of different α on epidemic spreading. $\alpha = 0, 1, 2, 3, 4, 5$, respectively. The greater the value of α , the greater the proportion of memory. $N = 10^4$, $L = 100$, $m = 1$, $0.0 < a < 0.8$, $\beta_I = 0.8$, $\beta_E = 0.08$, $\varepsilon_E = 0.08$, $\xi_I = 0.0004$, and the distribution of activity follows a power-law distribution with a power exponent of 2.8. The greater the value of α , the stronger the speed, and the scope of the epidemic spreading is inhibited.

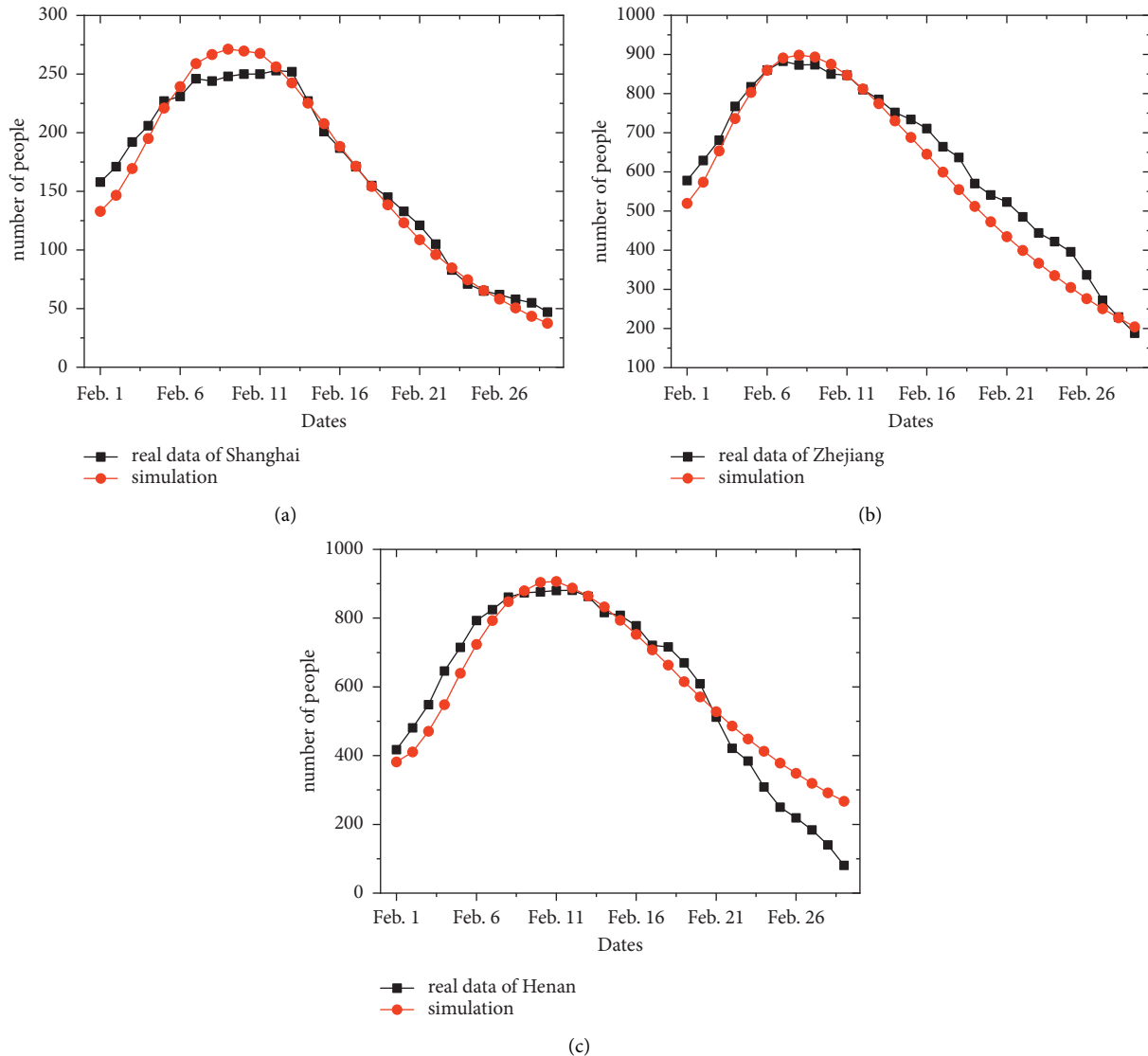


FIGURE 7: The simulated number of infected persons on the TSM model and the real number of confirmed cases in three regions in February, 2020. (a) Simulated and real data of Shanghai. (b) Simulated and real data of Zhejiang. (c) Simulated and real data of Henan. The propagation trend in time-varying networks and real networks is roughly the same. It is of great practical significance to study the propagation process on time-varying networks.

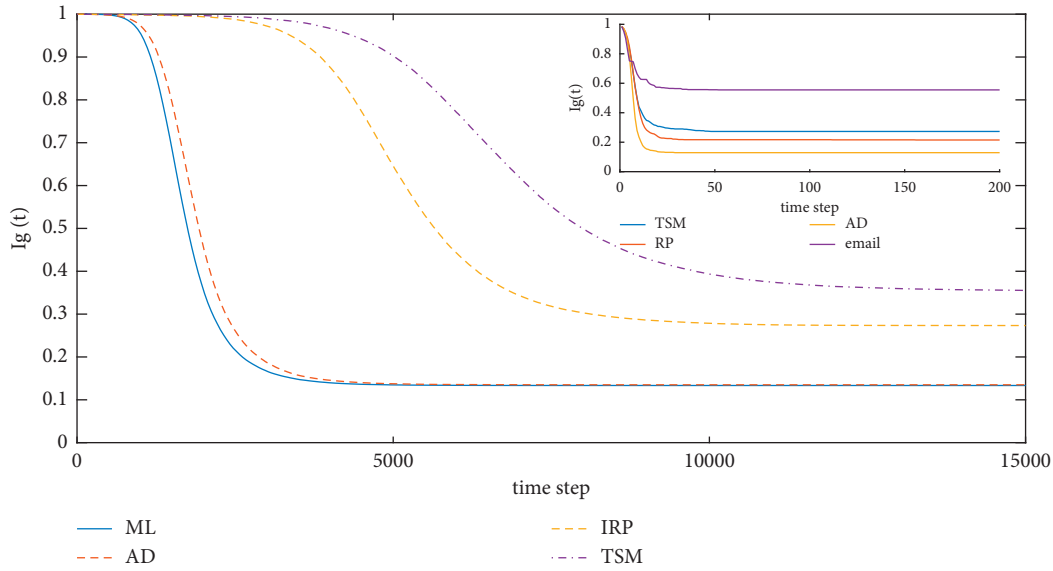


FIGURE 8: $I_g(t)$ over time during rumor propagation on the network of ML, AD, IRP, and TSM models. The rumor propagation speed in the AD model drops slightly, and there is basically no impact on the spreading ranges. The IRP model has an obvious constraint on rumor spreading. In the TSM model, the propagation speed and range are further controlled. Subgraph is the comparison of rumor propagation results between the simulated network and actual mail network. $N = 10^5$, $m = 1$, $\lambda = 1$, and $\mu = 0.6$, and the distribution of activity follows a power-law distribution with a power exponent of 2.8.

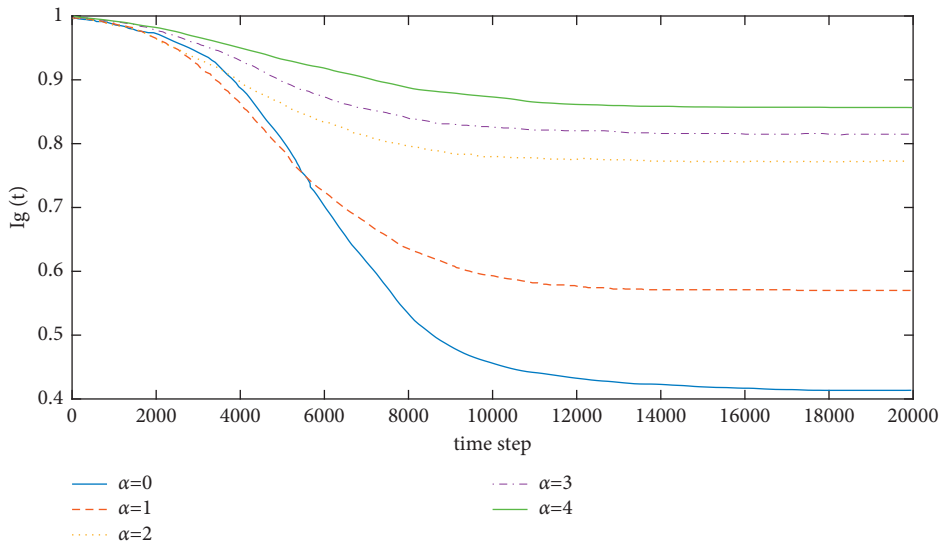


FIGURE 9: The influence of different α on the rumor propagation on the network. The inhibition effect of memory is more stronger. $\alpha = 0, 1, 2, 3, 4, 5$, respectively, $N = 10^5$, $m = 1$, $\lambda = 1$, and $\mu = 0.6$, and the distribution of activity follows a power-law distribution with a power exponent of 2.8.

inflow rate in the area k is φ_k , and the population outflow rate is ρ_k . $\delta < 0$ indicates that the outflow of population is greater than the inflow of population, while $\delta > 0$ indicates the opposite. Taking Shanghai as an example, the epidemic trend under different inflow and outflow rates is simulated, and the results are shown in Figure 11. It is found that, after the epidemic has been basically controlled, the impact of urban population decreasing on the epidemic situation is small. For cities with an influx of population, δ has little impact on the

epidemic when it is relatively small ($\delta \leq 0.14$). However, as δ is gradually increasing, the epidemic is getting serious. It will lead to the outbreak of the epidemic again in a short time when $\delta \geq 0.20$, so it is important to control the population flow during the return journey.

In order to study the influencing factors of work resumption, we analyze the comprehensive effects of the delay in returning and the inflow rate on the epidemic. It can be found from Figure 12(a) that returning to the city at a higher

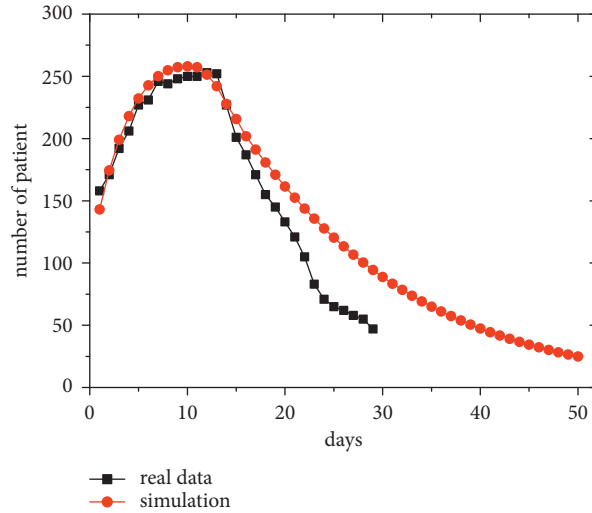


FIGURE 10: The simulated result of the time-varying interregional epidemic model and the real number of confirmed cases in Shanghai in February, 2020. The red curve indicates the actual number of COVID-19 patients per day. The black curve is simulated by the interregional model. In February, our simulation results fit well with the real results, and we also made predictions about the epidemic situation after February.

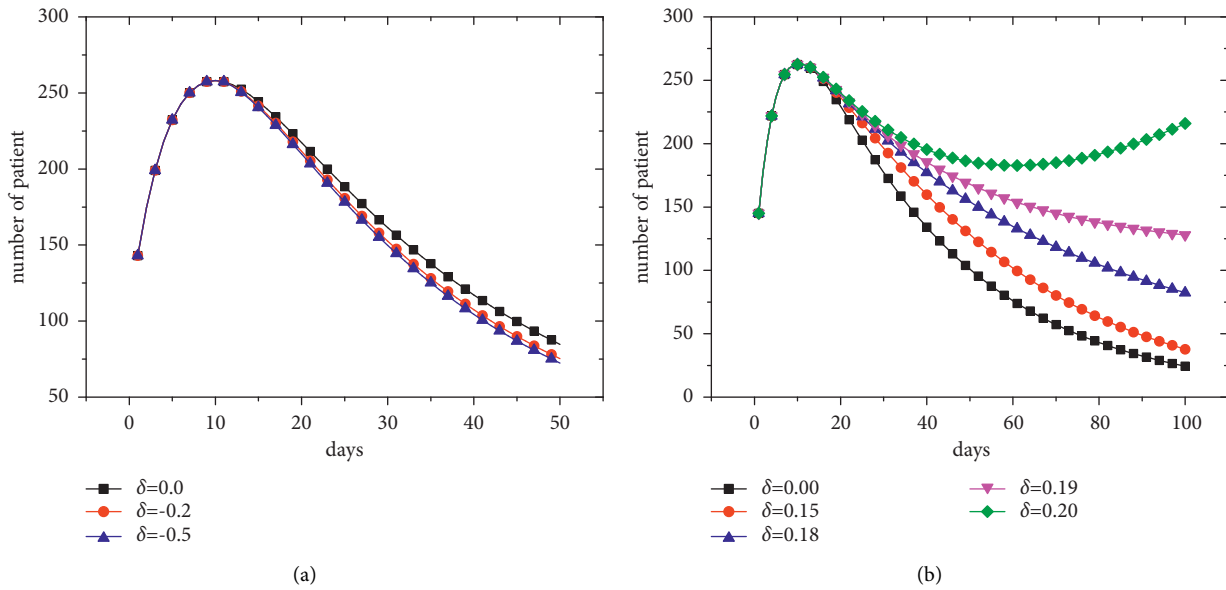


FIGURE 11: The impact of different rates on the epidemic. (a) The impact of different outflow rates on the epidemic when the outflow of population is greater. After the epidemic has been basically controlled, the impact of the urban population decreasing on the epidemic situation is small. (b) The impact of different inflow rates on the epidemic when the inflow of population is greater, and the x -axis is extended to 100 days to see the trend. δ has little impact on the epidemic when it is relatively small ($\delta \leq 0.14$). While varying δ from 0.14 to 0.20, the epidemic is getting serious. Also, it will lead to the outbreak of the epidemic again in a short time when $\delta \geq 0.20$.

rate of population inflow after 7 days can effectively reduce the epidemic. We further consider the impact of returning to the city in batches, and the result in Figure 12(a) indicates that returning to the city in batches can ensure the efficiency of resumption of work and avoid a great impact on the epidemic.

Due to the lack of reliable and efficient virus detection methods, human errors, and some unregulated population

inflow, it is impossible to effectively and accurately detect and isolate all infected people returning to work. Figure 12(b) shows the epidemic trend under two different migration policies ($c = 0$ and $c = 1$) with different δ . It is easy to find from Figure 12(b) that even if all infected persons are effectively isolated, it will be of little help to control the epidemic under the increasing inflow of population. The results show that when there are enough exposed persons,

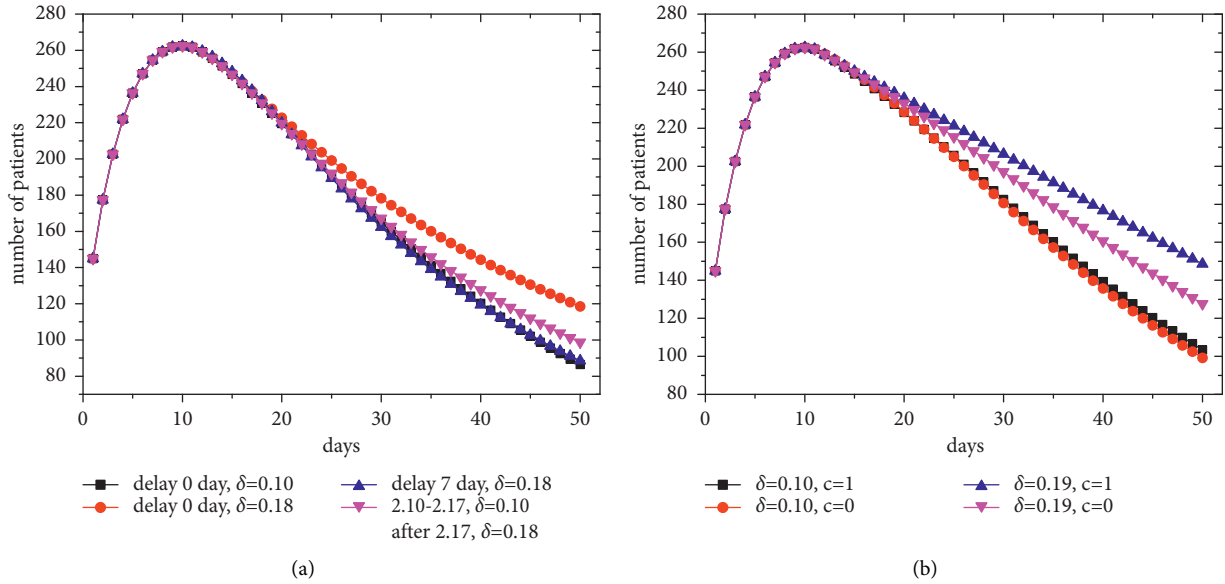


FIGURE 12: The trend of the epidemic under different conditions. (a) The relationship between postponing the resumption of work and increasing population inflow. Returning to the city at a higher rate of population inflow after 7 days can effectively reduce the epidemic, and returning to the city in batches can also ensure the efficiency of resumption of work and avoid a great impact on the epidemic. (b) The trend of the epidemic under different δ and c . When there are enough exposed persons, even if the infected people are completely isolated, the development of the epidemic cannot be effectively prevented.

even if the infected people are completely isolated, the development of the epidemic cannot be effectively prevented.

6. Conclusions

Considering that the spatiality and memory attribute of nodes are two important driving factors to promote the evolution of time-varying networks besides activity, we propose the TSM model by combining the distance and memory. If the cumulative interaction between two nodes is more or the two nodes are closer, the probability of establishing edges between them is greater. The proportion of the two driving factors in the model can be adjusted by parameters, so that the model can be adapted to a variety of practical time-varying networks.

In order to provide a reference for the spread and management of COVID-19, we apply the TSM model to pandemic spreading. The distance and intimacy between nodes simultaneously are driven by the network structure in the TSM model, so the epidemic spreading on the network is further reduced and slowed down in scope and speed. We also find that the memory is more important in inhibiting the spreading. Considering the requirements of epidemic prevention and management, we limit the active behaviors of nodes. The results show that when the activity of nodes decreases or the contacts created by active nodes per time step is reduced, the epidemic will be effectively controlled. In addition, increasing the recovery probability of infected people in each time step can greatly reduce the spread of the epidemic. We also verify the validity of the experiment on time-varying networks for epidemic spreading through real data, and the simulation

result is the same as the trend of the real data. We then find that our model is also applicable to rumor propagation.

Finally, in order to understand the impact of inter-regional population flow caused by work resumption on the epidemic, the model based on the population mobility rate is proposed. We determine the relevant parameters of the epidemiological characteristics of COVID-19 through relevant literature and daily cases and prove that the model and parameters can fit the real data well. Then, the effectiveness of different control measures is evaluated, such as changing the flow rate, delaying the time to return to work, and strengthening the screening of inflows. Furthermore, we analyze the relationship between return time and virus transmission and find that delaying return time can reduce the spread of the epidemic and an appropriate δ increasing with the return time can have a less impact on the epidemic. In order to take into account the economic development at the same time, we suggest that we should let workers return to work in cities while controlling the epidemic, and the strategy of resuming work in batches can be implemented.

In this paper, we focus on the research of modeling the temporal interpersonal contact network of infectious disease and reflecting COVID-19 dynamics, while the threshold of the spreading or other characteristics are important from the overall perspective of epidemic spreading. The stability or critical characteristics of the proposed epidemic model deserves further investigation in the future.

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 no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 11875133 and 11075057) and the National Key Research and Development Program of China (Grant no. 2018YFB2101302).

References

- [1] World Health Organization, *Coronavirus*, WHO, Geneva, Switzerland, 2020, <https://www.who.int/health-topics/coronavirus/enleadertwodots>.
- [2] S. Zhao, Z. Zhuang, J. Ran et al., “The association between domestic train transportation and novel coronavirus (2019-ncov) outbreak in China from 2019 to 2020: a data-driven correlational report,” *Travel Medicine and Infectious Disease*, vol. 33, Article ID 101568, 2020.
- [3] C. Huang, Y. Wang, X. Li et al., “Clinical features of patients infected with 2019 novel coronavirus in wuhan, China,” *The Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.
- [4] J. F.-W. Chan, S. Yuan, K.-H. Kok et al., “A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster,” *The Lancet*, vol. 395, no. 10223, pp. 514–523, 2020.
- [5] Q. Li, X. Guan, P. Wu et al., “Early transmission dynamics in wuhan, China, of novel coronavirus-infected pneumonia,” *New England Journal of Medicine*, vol. 382, no. 13, pp. 1199–1207, 2020.
- [6] Y. Yin, J. Gao, B. F. Jones, and D. Wang, “Coevolution of policy and science during the pandemic,” *Science*, vol. 371, no. 6525, pp. 128–130, 2021.
- [7] X. Hao, S. Cheng, D. Wu, T. Wu, X. Lin, and C. Wang, “Reconstruction of the full transmission dynamics of covid-19 in wuhan,” *Nature*, vol. 584, no. 7821, pp. 420–424, 2020.
- [8] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, and N. A. Christakis, “Population flow drives spatio-temporal distribution of covid-19 in China,” *Nature*, vol. 582, no. 7812, pp. 389–394, 2020.
- [9] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez et al., “The effect of human mobility and control measures on the covid-19 epidemic in China,” *Science*, vol. 368, no. 6490, pp. 493–497, 2020.
- [10] A. Arenas, W. Cota, J. Gómez-Gardeñes et al., “Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions,” *Physical Review X*, vol. 10, no. 4, Article ID 41055, 2020.
- [11] L. Basnarkov, Epidemic Spreading Model of Covid-19, 2020.
- [12] L. F. S. Scabini, L. C. Ribas, M. B. Neiva, A. G. B. Junior, A. J. F. Farfán, and O. M. Bruno, “Social interaction layers in complex networks for the dynamical epidemic modeling of covid-19 in Brazil,” *Physica A*, vol. 564, Article ID 125498, 2021.
- [13] E. Montes-Orozco, R.-A. Mora-Gutiérrez, S.-G. De-Los-Cobos-Silva et al., “Identification of covid-19 spreaders using multiplex networks approach,” *IEEE Access*, vol. 8, pp. 122874–122883, 2020.
- [14] Z. Wang, C. Xia, Z. Chen, and G. Chen, “Epidemic propagation with positive and negative preventive information in multiplex networks,” *IEEE transactions on cybernetics*, vol. 51, no. 3, pp. 1454–1462, 2020.
- [15] Z. Wang and C. Xia, “Co-evolution spreading of multiple information and epidemics on two-layered networks under the influence of mass media,” *Nonlinear Dynamics*, vol. 102, no. 4, pp. 3039–3052, 2020.
- [16] P. Holme and J. Saramäki, “Temporal networks,” *Physics Reports*, vol. 519, no. 3, pp. 97–125, 2012.
- [17] L. E. C. Rocha, “Dynamics of air transport networks: a review from a complex systems perspective,” *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 469–478, 2017.
- [18] S. Sarkar, S. Sikdar, S. Bhowmick, and A. Mukherjee, “Using core-periphery structure to predict high centrality nodes in time-varying networks,” *Data Mining and Knowledge Discovery*, vol. 32, no. 5, pp. 1368–1396, 2018.
- [19] Q. Chen, J. H. Qian, L. Zhu, and D. D. Han, “Optimal transport in time-varying small-world networks,” *Physical Review*, vol. 93, no. 3, Article ID 32321, 2016.
- [20] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, “Activity driven modeling of time varying networks,” *Scientific Reports*, vol. 2, no. 1, p. 469, 2012.
- [21] 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 66130, 2004.
- [22] L. Alessandretti, K. Sun, A. Baronchelli, and N. Perra, “Random walks on activity-driven networks with attractiveness,” *Physical Review*, vol. 95, no. 5, Article ID 52318, 2017.
- [23] A. Moinet, A. Barrat, and R. Pastor-Satorras, “Generalized voterlike model on activity-driven networks with attractiveness,” *Physical Review*, vol. 98, no. 2, Article ID 22303, 2018.
- [24] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, “Modeling human dynamics of face-to-face interaction networks,” *Physical Review Letters*, vol. 110, no. 16, Article ID 168701, 2013.
- [25] T. Aoki, L. E. Rocha, and T. Gross, “Temporal and structural heterogeneities emerging in adaptive temporal networks,” *Physical Review*, vol. 93, no. 4, Article ID 40301, 2016.
- [26] J. Wang, J. Xu, Y. Liu, and W. Deng, “AST: activity-security-trust driven modeling of time varying networks,” *Scientific Reports*, vol. 6, no. 1, p. 21352, 2016.
- [27] H. Dun, Y. Shuting, H. She, Q. Lingfei, and A. B. Chris, “Research on how the difference of personal propagation ability influences the epidemic spreading in activity-driven network,” *Physica A: Statistical Mechanics and Its Applications*, vol. 514, pp. 311–318, 2019.
- [28] X. An, L. Ding, and P. Hu, “Information propagation with individual attention-decay effect on activity-driven networks,” *Physica A: Statistical Mechanics and Its Applications*, vol. 556, Article ID 124815, 2020.
- [29] L. Han, Z. Lin, M. Tang, J. Zhou, Y. Zou, and S. Guan, “Impact of contact preference on social contagions on complex networks,” *Physical Review*, vol. 101, no. 4, Article ID 42308, 2020.
- [30] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, “Modeling echo chambers and polarization dynamics in social networks,” *Physical Review Letters*, vol. 124, no. 4, Article ID 48301, 2020.
- [31] Y. Cheng, C. Liu, and F. Ding, “Dynamic analysis of rumor spreading model for considering active network nodes and nonlinear spreading rate,” *Physica A: Statistical Mechanics and its Applications*, vol. 506, pp. 24–35, 2018.
- [32] H. Zhu, J. Ma, and S. Li, “Effects of online and offline interaction on rumor propagation in activity-driven networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 525, pp. 1124–1135, 2019.

- [33] H. Kim, M. Ha, and H. Jeong, "Impact of temporal connectivity patterns on epidemic process," *The European Physical Journal B*, vol. 92, no. 7, p. 161, 2019.
- [34] M. Tizzani, S. Lenti, E. Ubaldi, A. Vezzani, C. Castellano, and R. Burioni, "Epidemic spreading and aging in temporal networks with memory," *Physical Review E*, vol. 98, no. 6, Article ID 62315, 2018.
- [35] C. L. Vestergaard, M. Génois, and A. Barrat, "How memory generates heterogeneous dynamics in temporal networks," *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 90, no. 4, Article ID 42805, 2014.
- [36] Y. Zhou, Z. Yue, Q.-H. Liu, and W. Wang, "Social reinforcement inducing discontinuous spreading in complex networks," *EPL (Europhysics Letters)*, vol. 128, no. 6, p. 68002, 2020.
- [37] O. E. Williams, F. Lillo, and V. Latora, "Effects of memory on spreading processes in non-markovian temporal networks," *New Journal of Physics*, vol. 21, no. 4, Article ID 43028, 2019.
- [38] H. Kim, M. Ha, and H. Jeong, "Scaling properties in time-varying networks with memory," *The European Physical Journal B*, vol. 88, no. 12, p. 315, 2015.
- [39] H. Kim, M. Ha, and H. Jeong, "Dynamic topologies of activity-driven temporal networks with memory," *Physical Review*, vol. 97, no. 6, Article ID 62148, 2018.
- [40] M. Karsai, N. Perra, and A. Vespignani, "Time varying networks and the weakness of strong ties," *Scientific Reports*, vol. 4, no. 1, p. 4001, 2014.
- [41] A. B. Gumel, S. Ruan, T. Day et al., "Modelling strategies for controlling sars outbreaks," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 271, no. 1554, pp. 2223–2232, 2004.
- [42] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, 2020.
- [43] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, China: a modelling study," *The Lancet*, vol. 395, no. 10225, pp. 689–697, 2020.
- [44] J. Wallinga and M. Lipsitch, "How generation intervals shape the relationship between growth rates and reproductive numbers," *Proceedings of the Royal Society B: Biological Sciences*, vol. 274, no. 1609, pp. 599–604, 2007.

Research Article

Global Dynamics of an SIS Model on Metapopulation Networks with Demographics

Maoxing Liu ¹, Xinjie Fu ¹, Jie Zhang ¹ and Donghua Zhao ²

¹Department of Mathematics, North University of China, Taiyuan 030051, Shanxi, China

²School of Mathematical Science, Fudan University, Shanghai 200433, China

Correspondence should be addressed to Maoxing Liu; liumxsx@gmail.com

Received 16 August 2020; Revised 13 January 2021; Accepted 30 August 2021; Published 20 September 2021

Academic Editor: Sergio Gomez

Copyright © 2021 Maoxing Liu 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.

In this paper, we propose a susceptible-infected-susceptible (SIS) epidemic model with demographics on heterogeneous metapopulation networks. We analytically derive the basic reproduction number, which determines not only the existence of endemic equilibrium but also the global dynamics of the model. The model always has the disease-free equilibrium, which is globally asymptotically stable when the basic reproduction number is less than unity and otherwise unstable. We also provide sufficient conditions on the global stability of the unique endemic equilibrium. Numerical simulations are performed to illustrate the theoretical results and the effects of the connectivity and diffusion. Furthermore, we find that diffusion rates play an active role in controlling the spread of infectious diseases.

1. Introduction

In real life, the spread of infectious diseases is always affected by the flow of population. In order to study the impact of demographics and the flow of population, we will focus on an epidemic model with demographics on metapopulation networks. As Masuda [1] mentioned, underlying contact networks are considered to be static on the time scale of epidemics, whereas underlying humans' social networks of prevailing infectious diseases such as influenza are presumably dynamic even during one day. The dynamics of networks are induced by diffusion of individuals among residences, workplaces, places for social activities, and so on. As a simple framework within which the role of spatial processes in disease transmission can be examined, metapopulation models have been introduced to describe epidemics and ecological invasion in such a situation [1–3]. A node in such a model represents a metapopulation or a habitat and not an individual. A link represents a physical pathway connecting a pair of metapopulations. Individuals travel from one node to another. Simultaneously, interactions between individuals, such as infection, can occur in each metapopulation.

In recent years, the effect of diffusion on disease spread has caught the attention of many researchers [4–8]. Luca et al. [9] considered that metapopulation epidemic models can describe the spatial spread of an infectious disease through a spatially structured host population [10, 11]. These models consist of patches or subpopulations of the system, connected through a coupling process generally characterizing hosts' mobility [9, 12]. When the diffusion was first considered, the reaction-diffusion equation was regarded as a two-step process, and the reaction-diffusion was assumed to occur simultaneously to obtain the reaction-diffusion processes. Saldaña [13] derived continuous-time equations governing the limit dynamics of discrete-time reaction-diffusion processes defined on heterogeneous metapopulations, and the spread of infectious diseases under two different transmission mechanisms was considered. Juher et al. [14] proposed a system of continuous-time equations to analyze the spread of infectious diseases, which is based on the limited transmission and nonlimited transmission of diseases. Compared with reference [13], the authors studied the stability of endemic equilibrium in uncorrelated networks with nonlimited transmission. Therefore, a certain amount of diffusion is needed for the spread of diseases in

metapopulation situation. After that, Masuda [1] investigated the effect of diffusion rate on the spread of epidemics in metapopulation networks. It can be seen from the above work that diffusion can change the epidemic threshold in many heterogeneous networks [1].

Based on previous work, we will investigate an epidemic model with demographics in metapopulation networks. Here the structure of patches (nodes) is made up of the connectivity (degree) distribution. Moreover, each patch contains two types of individuals: susceptible and infected individuals. Within each patch, transmission or recovery (reaction processes) occurred between individuals of different types. The diffusion of individuals takes place at once among patches (diffusion process) at constant rates. In fact, reaction and diffusion processes can be considered to take place simultaneously. With this idea, many epidemic models have been proposed to describe the dynamics of disease spread among patches. The model of this paper is a description of an SIS metapopulation model at the node level with birth and death. Based on the model in [1], we consider the model with node-based birth and death. Wang et al. [15] made use of a mean-field approximation, and this model has birth and death, but its birth and death do not change with the change of degree. In some models, in order to facilitate calculation, the elements of the connectivity matrix A are given by $a_{ii'} = iP(i')/\langle i \rangle$. In this paper, we will analyze the model without using the mean-field approximation and transforming the elements of the adjacency matrix, and we study it in an arbitrary network.

The remaining of this paper is organized as follows. In Section 2, we firstly present the model and then study the existence of equilibria and their stability. There exist threshold dynamics that are determined by the basic reproduction number R_0 . The theoretical results and the effects of connectivity and diffusion are demonstrated by numerical simulations in Section 3. The paper concludes with a summary and a brief discussion in last section.

2. Main Results

2.1. The Model. For a metapopulation network, each node stands for a metapopulation or patch. Let N be the number of nodes; the $N \times N$ adjacency matrix, denoted by A , is defined by

$$A_{ji} = \begin{cases} 0, & \text{if } j = i \text{ or } j \text{ and } i \text{ are not adjacent when } j \neq i, \\ 1, & \text{if } j \neq i \text{ and } j \text{ and } i \text{ are adjacent.} \end{cases} \quad (1)$$

Clearly, A is symmetric. The model is an extension of the following one mentioned in Masuda [1]:

$$\begin{cases} \frac{d\rho_{S,i}}{dt} = -\beta\rho_{S,i}\rho_{I,i} + \gamma\rho_{I,i} - D_S\rho_{S,i} + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j}, \\ \frac{d\rho_{I,i}}{dt} = \beta\rho_{S,i}\rho_{I,i} - \gamma\rho_{I,i} - D_I\rho_{I,i} + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}, \end{cases} \quad (2)$$

where $\rho_{S,i}(t)$ and $\rho_{I,i}(t)$ represent the number of susceptible and infected individuals at node i and at time t , respectively, D_S and D_I are the diffusion rates for the susceptible and infected individuals, respectively, β is the transmission rate, γ is the recovery rate, and k_i is the degree of the node i . For (2), it is shown that diffusion increases the epidemic threshold of the SIS dynamics in arbitrary heterogeneous networks.

In this paper, we incorporate demographics into (2), and the model is as follows:

$$\begin{cases} \frac{d\rho_{S,i}}{dt} = b_i - \beta\rho_{S,i}\rho_{I,i} + \gamma\rho_{I,i} - D_S\rho_{S,i} + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j} - \mu\rho_{S,i}, \\ \frac{d\rho_{I,i}}{dt} = \beta\rho_{S,i}\rho_{I,i} - \gamma\rho_{I,i} - D_I\rho_{I,i} + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j} - \mu\rho_{I,i}, \end{cases} \quad (3)$$

where b_i is the birth rate at node i and μ is the natural death rate of individuals (whose meaning is different from that in (2)).

It is easy to see that (3) is well posed, that is, for any given nonnegative initial condition, it has a unique and global solution through it. Moreover, the solution is also non-negative. Let $\bar{b} = \sum_{i=1}^N b_i$ and $H = \sum_{i=1}^N (\rho_{S,i} + \rho_{I,i})$. Adding up all equations of (3) yields

$$\frac{dH(t)}{dt} = \bar{b} - \mu H(t), \quad (4)$$

which implies that $\limsup_{t \rightarrow \infty} H(t) \leq (\bar{b}/\mu)$. Thus, the feasible region of (3) can be chosen as

$$\Gamma = \left\{ (\rho_{S,1}, \dots, \rho_{S,N}, \rho_{I,1}, \dots, \rho_{I,N}) \in \mathbb{R}_+^{2N} \mid \sum_{i=1}^N (\rho_{S,i} + \rho_{I,i}) \leq \frac{\bar{b}}{\mu} \right\}. \quad (5)$$

It can be verified that Γ is attractive and positively invariant with respect to (3) (see [16]). Let $\dot{\Gamma}$ denote the interior of Γ and $\partial\Gamma$ denote the boundary of Γ .

2.2. Existence of Disease-Free Equilibria. We start the analysis of (3) with the existence of disease-free equilibria. At a disease-free equilibrium of (3), we have the following linear system [17]:

$$b_i - D_S\rho_{S,i} + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j} - \mu\rho_{S,i} = 0, \quad i = 1, 2, \dots, N, \quad (6)$$

which can be written in the matrix form

$$\bar{L}S = \bar{B}, \quad (7)$$

where

$$\bar{L} = \begin{pmatrix} D_S + \mu & -D_S \frac{A_{21}}{k_2} & \dots & -D_S \frac{A_{N1}}{k_N} \\ -D_S \frac{A_{12}}{k_1} & D_S + \mu & \dots & -D_S \frac{A_{N2}}{k_N} \\ \vdots & \vdots & \ddots & \vdots \\ -D_S \frac{A_{1N}}{k_1} & -D_S \frac{A_{2N}}{k_2} & \dots & D_S + \mu \end{pmatrix},$$

$$\begin{aligned} S &= (\rho_{S,1}, \rho_{S,2}, \dots, \rho_{S,N})^T, \\ \bar{B} &= (b_1, b_2, \dots, b_N)^T. \end{aligned} \quad (8)$$

As \bar{L} is a column diagonal dominant matrix with positive diagonal elements and negative off-diagonal elements, \bar{L} is nonsingular and any element of \bar{L}^{-1} is positive [18]. Thus, (6) has a unique positive solution, which has the following result.

Proposition 1. *System (3) always has a unique disease-free equilibrium, denoted by $E_0 = (\rho_{S,1}^0, \dots, \rho_{S,N}^0, 0, \dots, 0)$.*

Since the adjacency matrix A_{ji} is randomly given, the disease-free equilibrium E_0 given in Proposition 1 does not have an explicit expression. Thus, we calculate it in specific network: globally coupled network. In a globally coupled network,

$$A = \begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{pmatrix}. \quad (9)$$

Then,

$$\bar{L} = \begin{pmatrix} D_S + \mu & -\frac{D_S}{N-1} & \dots & -\frac{D_S}{N-1} \\ -\frac{D_S}{N-1} & D_S + \mu & \dots & -\frac{D_S}{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{D_S}{N-1} & -\frac{D_S}{N-1} & \dots & D_S + \mu \end{pmatrix}. \quad (10)$$

By complex calculations, we get

$$\bar{L}^{-1} = \begin{pmatrix} \frac{(N-1)\mu + D_S}{N\mu D_S + \mu^2(N-1)} & \frac{D_S}{N\mu D_S + \mu^2(N-1)} & \dots & \frac{D_S}{N\mu D_S + \mu^2(N-1)} \\ \frac{D_S}{N\mu D_S + \mu^2(N-1)} & \frac{(N-1)\mu + D_S}{N\mu D_S + \mu^2(N-1)} & \dots & \frac{D_S}{N\mu D_S + \mu^2(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{D_S}{N\mu D_S + \mu^2(N-1)} & \frac{D_S}{N\mu D_S + \mu^2(N-1)} & \dots & \frac{(N-1)\mu + D_S}{N\mu D_S + \mu^2(N-1)} \end{pmatrix}, \quad (11)$$

and hence

$$\rho_{S,i}^0 = \frac{[(N-1)\mu + D_S]b_i + D_S(-b_i + \sum_{j=1}^N b_j)}{N\mu D_S + \mu^2(N-1)}, \quad \text{for } i = 1, 2, \dots, N. \quad (12)$$

In particular, if $b_1 = b_2 = \dots = b_N = b$, then $\rho_{S,i}^0 = (\mu/b)$ and the disease-free equilibrium is $E_0 = (\underbrace{(\mu/b), \dots, (\mu/b)}_N, 0, \dots, 0)$.

2.3. Stability of the Disease-Free Equilibrium. Before considering the stability of the disease-free equilibrium of (3), we introduce the basic reproduction number.

Linearizing (3) at the disease-free equilibrium E_0 , we can get

$$\frac{dI(t)}{dt} = MI(t), \quad (13)$$

where $I(t) = (\rho_{I,1}(t), \rho_{I,2}(t), \dots, \rho_{I,N}(t))^T$ and

$$M = \begin{pmatrix} \beta\rho_{S,1}^0 - \gamma - \mu - D_I & D_I \frac{A_{21}}{k_2} & \dots & D_I \frac{A_{N1}}{k_N} \\ D_I \frac{A_{12}}{k_1} & \beta\rho_{S,2}^0 - \gamma - \mu - D_I & \dots & D_I \frac{A_{N2}}{k_N} \\ \vdots & \vdots & \ddots & \vdots \\ D_I \frac{A_{1N}}{k_1} & D_I \frac{A_{2N}}{k_2} & \dots & \beta\rho_{S,N}^0 - \gamma - \mu - D_I \end{pmatrix}. \quad (14)$$

Then, $s(M)$ is a simple eigenvalue of M with a positive eigenvector. Let $F_i(I)$ denote the rate of appearance of new infections in node i , $V_i^+(I)$ be the transfer rate of infected

individuals into node i by all other means, and $V_i^-(I)$ be the transfer rate of infected individuals out of node i . Then, we can choose

$$F = \begin{pmatrix} \beta\rho_{S,1}^0 & 0 & \cdots & 0 \\ 0 & \beta\rho_{S,2}^0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta\rho_{S,N}^0 \end{pmatrix}, \quad (15)$$

$$V = \begin{pmatrix} V_1^-(I) - V_1^+(I) \\ V_2^-(I) - V_2^+(I) \\ \vdots \\ V_N^-(I) - V_N^+(I) \end{pmatrix} = \begin{pmatrix} \mu + D_I + \gamma & -D_I \frac{A_{21}}{k_2} & \cdots & -D_I \frac{A_{N1}}{k_N} \\ -D_I \frac{A_{12}}{k_1} & \mu + D_I + \gamma & \cdots & -D_I \frac{A_{N2}}{k_N} \\ \vdots & \vdots & \ddots & \vdots \\ -D_I \frac{A_{1N}}{k_1} & -D_I \frac{A_{2N}}{k_2} & \cdots & \mu + D_I + \gamma \end{pmatrix}. \quad (16)$$

We call FV^{-1} the next generation matrix for (3), and the basic reproduction number R_0 is defined as

$$R_0 = \rho(FV^{-1}), \quad (17)$$

where $\rho(FV^{-1})$ denotes the spectral radius of the matrix FV^{-1} (see [19] for details). Note that R_0 is simply the product of the infection rate and the mean duration of the infection [15]. In a globally coupled network, the basic reproduction number is $R_0 = (b\beta/\mu(\gamma + \mu))$. Refer to [19] for Theorem 1.

Theorem 1. *The following statements on (3) are true.*

- (i) $R_0 > 1$ if and only if $s(M) > 0$ and $R_0 < 1$ if and only if $s(M) < 0$.
- (ii) If $R_0 < 1$, then the disease-free equilibrium E_0 is locally asymptotically stable while it is unstable if $R_0 > 1$.

Now, we establish the global stability of the disease-free equilibrium.

Theorem 2. *Assume $R_0 < 1$. Suppose that A is irreducible. Then, the disease-free equilibrium E_0 of (3) is globally asymptotically stable in R_+^{2N} .*

Proof. As Γ is an attracting and positively invariant set, we only need to show the global stability of E_0 in Γ .

Let F and V be given by (15) and (16), respectively. Since all off-diagonal entries of V are nonpositive and the sum of the entries in each column of V is positive, V is a nonsingular M -matrix. Also, V^{-1} is irreducible. By the Perron–Frobenius theorem ([18], p. 27), the nonnegative irreducible matrix $V^{-1}F$ has a positive left eigenvector (w_1, w_2, \dots, w_N)

corresponding to the eigenvalue $\rho(V^{-1}F)$. Since F is a diagonal matrix, $\rho(V^{-1}F) = \rho(FV^{-1})$. Consequently, we have

$$(w_1, w_2, \dots, w_N)V^{-1}F = R_0(w_1, w_2, \dots, w_N), \quad (18)$$

or

$$\frac{1}{R_0}(w_1, w_2, \dots, w_N) = (w_1, w_2, \dots, w_N)F^{-1}V. \quad (19)$$

For $i = 1, \dots, N$, denote $c_i = (w_i/\beta\rho_{S,i}^0) > 0$ and $I = (\rho_{I,1}, \rho_{I,2}, \dots, \rho_{I,N})^T$. Consider the Lyapunov function

$$L = \sum_{i=1}^N c_i \rho_{I,i}, \quad (20)$$

and the derivative of L along the solutions of (3) is

$$\begin{aligned} \frac{dL}{dt}|_{(2,2)} &= \sum_{i=1}^N c_i \left(\beta\rho_{S,i}\rho_{I,i} - \gamma\rho_{I,i} - D_I\rho_{I,i} + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j} - \mu\rho_{I,i} \right) \\ &\leq \sum_{i=1}^N c_i (\beta\rho_{S,i}^0 - \gamma - \mu - D_I) \rho_{I,i} + D_I \sum_{j \neq i}^N \frac{A_{ji}}{k_j} \rho_{I,j} \\ &= \left(\frac{w_1}{\beta\rho_{S,1}^0}, \frac{w_2}{\beta\rho_{S,2}^0}, \dots, \frac{w_N}{\beta\rho_{S,N}^0} \right) (F - V)I \\ &= (w_1, w_2, \dots, w_N) (1 - F^{-1}V)I \\ &= (w_1, w_2, \dots, w_N) \left(1 - \frac{1}{R_0} \right) I \leq 0. \end{aligned} \quad (21)$$

Since $c_i > 0$ for all i , $(dL/dt)|_{(2.2)} = 0$ implies that either $\rho_{S,i} = \rho_{S,i}^0$ or $\rho_{I,i} = 0$ for any $1 \leq i \leq N$. When $\rho_{S,i} = \rho_{S,i}^0$, it follows from the first equation of (3) that

$$0 = \frac{d\rho_{S,i}^0}{dt} = b_i - \beta\rho_{S,i}^0\rho_{I,i} + \gamma\rho_{I,i} - D_S\rho_{S,i}^0 + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j}^0 - \mu\rho_{S,i}^0. \quad (22)$$

Comparing this with (6) gives $\rho_{I,i} = 0$. Thus, we have shown that $(dL/dt)|_{(2.2)} = 0$ which implies that $\rho_{I,i} = 0$ for all i . By LaSalle's invariance principle, we see that E_0 is globally asymptotically stable in Γ and hence in R_+^{2N} .

The following result tells us that when $R_0 > 1$, the disease is persistent. \square

Proposition 2 (see [17]). *Suppose that A is irreducible. If $R_0 > 1$, then system (3) is uniformly persistent and there exists an endemic equilibrium E^* in $\mathring{\Gamma}$.*

The proof of Proposition 2 is similar to the proof of Proposition 3.3 in [20]. Using uniform persistence of (3) and uniform boundedness of solutions in $\mathring{\Gamma}$ we can obtain the existence of an equilibrium of (3) in $\mathring{\Gamma}$ (see Theorem 2.8.6 in [21]). In Proposition 2, the assumption that the adjacency matrix $A = (A_{ij})$ is irreducible is necessary. If $A = 0$, then system (3) has an asymptotically stable boundary equilibrium when $R_0 > 1$, and thus the system is not persistent [22].

2.4. The Global Stability of the Endemic Equilibrium. In this section, we provide sufficient conditions on the global stability of the endemic equilibrium. The method of proof is the graph-theoretical approach developed in [23–25].

Theorem 3. *Assume that $R_0 > 1$. Let $E^* = (\rho_{S,1}^*, \dots, \rho_{S,N}^*, \rho_{I,1}^*, \dots, \rho_{I,N}^*)$ be an endemic equilibrium of (3). If A is irreducible and there exists $\lambda > 0$ such that $D_S \sum_{j=1}^N (A_{ji}/k_j) \rho_{S,j}^* = \lambda D_I \sum_{j=1}^N (A_{ji}/k_j) \rho_{I,j}^*$ for all $1 \leq i, j \leq N$, then the endemic equilibrium E^* is globally asymptotically stable in $\mathring{\Gamma}$, and hence (3) only has a unique endemic equilibrium.*

Proof. For each $1 \leq i \leq N$, set

$$V_i(\rho_{S,i}, \rho_{I,i}) = \rho_{S,i} - \rho_{S,i}^* - \rho_{S,i}^* \ln \frac{\rho_{S,i}}{\rho_{S,i}^*} + \rho_{I,i} - \rho_{I,i}^* - \rho_{I,i}^* \ln \frac{\rho_{I,i}}{\rho_{I,i}^*}. \quad (23)$$

Noting

$$\begin{aligned} D_S \rho_{S,i}^* &= b_i - \beta \rho_{S,i}^* \rho_{I,i}^* + \gamma \rho_{I,i}^* + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j}^* - \mu \rho_{S,i}^*, \\ (D_I + \mu) \rho_{I,i}^* &= \beta \rho_{S,i}^* \rho_{I,i}^* + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}^* - \gamma \rho_{I,i}^*, \end{aligned} \quad (24)$$

we calculate the derivative of V_i along solutions of (3) as

$$\begin{aligned} \frac{dV_i}{dt}|_{(2.2)} &= b_i - D_S \rho_{S,i} + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j} - \mu \rho_{S,i} - b_i \frac{\rho_{S,i}}{\rho_{S,i}^*} + \beta \rho_{I,i} \rho_{S,i}^* \\ &\quad - \gamma \rho_{I,i} \frac{\rho_{S,i}}{\rho_{S,i}^*} + D_S \rho_{S,i}^* - D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j} \frac{\rho_{S,i}}{\rho_{S,i}^*} + \mu \rho_{S,i}^* - D_I \rho_{I,i} \\ &\quad + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j} - \mu \rho_{I,i} - \beta \rho_{S,i} \rho_{I,i}^* + \gamma \rho_{I,i}^* + D_I \rho_{I,i}^* + \mu \rho_{I,i}^* - D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j} \frac{\rho_{I,i}}{\rho_{I,i}^*} \\ &= b_i \left[1 - \frac{\rho_{S,i}}{\rho_{S,i}^*} + \ln \frac{\rho_{S,i}}{\rho_{S,i}^*} + 1 - \frac{\rho_{S,i}}{\rho_{S,i}} + \ln \frac{\rho_{S,i}^*}{\rho_{S,i}} \right] \\ &\quad + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j}^* \left[1 - \frac{\rho_{S,i} \rho_{S,j}}{\rho_{S,i} \rho_{S,j}^*} + \ln \frac{\rho_{S,i} \rho_{S,j}}{\rho_{S,i} \rho_{S,j}^*} \right] \\ &\quad + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j}^* \left[\frac{\rho_{S,j}}{\rho_{S,j}^*} + \ln \frac{\rho_{S,j}}{\rho_{S,j}^*} - \frac{\rho_{S,i}}{\rho_{S,i}^*} - \ln \frac{\rho_{S,i}}{\rho_{S,i}^*} \right] \\ &\quad + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}^* \left[1 - \frac{\rho_{I,i} \rho_{I,j}}{\rho_{I,i} \rho_{I,j}^*} + \ln \frac{\rho_{I,i} \rho_{I,j}}{\rho_{I,i} \rho_{I,j}^*} \right] \end{aligned}$$

$$\begin{aligned}
& + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}^* \left[\frac{\rho_{I,j}}{\rho_{I,j}^*} + \ln \frac{\rho_{I,j}^*}{\rho_{I,j}} - \frac{\rho_{I,i}}{\rho_{I,i}^*} - \ln \frac{\rho_{I,i}^*}{\rho_{I,i}} \right] \\
& + \gamma \rho_{I,i}^* \left[1 - \frac{\rho_{S,i}^* \rho_{I,i}}{\rho_{S,i} \rho_{I,i}^*} + \ln \frac{\rho_{S,i}^* \rho_{I,i}}{\rho_{S,i} \rho_{I,i}^*} \right] + \gamma \rho_{I,i}^* \left[\ln \frac{\rho_{I,i}^*}{\rho_{I,i}} - \ln \frac{\rho_{S,i}^*}{\rho_{S,i}} \right] \\
& + \gamma \rho_{I,i} \left[1 - \frac{\rho_{I,i}^* \rho_{S,i}}{\rho_{I,i} \rho_{S,i}^*} + \ln \frac{\rho_{I,i}^* \rho_{S,i}}{\rho_{I,i} \rho_{S,i}^*} \right] + \gamma \rho_{I,i} \left[\ln \frac{\rho_{I,i}}{\rho_{I,i}^*} - \ln \frac{\rho_{S,i}}{\rho_{S,i}^*} \right] \\
\leq & D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j}^* \left[\frac{\rho_{S,j}}{\rho_{S,j}^*} + \ln \frac{\rho_{S,j}^*}{\rho_{S,j}} - \frac{\rho_{S,i}}{\rho_{S,i}^*} - \ln \frac{\rho_{S,i}^*}{\rho_{S,i}} \right] \\
& + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}^* \left[\frac{\rho_{I,j}}{\rho_{I,j}^*} + \ln \frac{\rho_{I,j}^*}{\rho_{I,j}} - \frac{\rho_{I,i}}{\rho_{I,i}^*} - \ln \frac{\rho_{I,i}^*}{\rho_{I,i}} \right] \\
& + \gamma \rho_{I,i}^* \left[\ln \frac{\rho_{I,i}^*}{\rho_{I,i}} - \ln \frac{\rho_{S,i}^*}{\rho_{S,i}} \right] + \gamma \rho_{I,i} \left[\ln \frac{\rho_{I,i}}{\rho_{I,i}^*} - \ln \frac{\rho_{S,i}}{\rho_{S,i}^*} \right] \\
= & D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j}^* \left[\frac{\rho_{S,j}}{\rho_{S,j}^*} + \ln \frac{\rho_{S,j}^*}{\rho_{S,j}} - \frac{\rho_{S,i}}{\rho_{S,i}^*} - \ln \frac{\rho_{S,i}^*}{\rho_{S,i}} \right] \\
& + D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}^* \left[\frac{\rho_{I,j}}{\rho_{I,j}^*} + \ln \frac{\rho_{I,j}^*}{\rho_{I,j}} - \frac{\rho_{I,i}}{\rho_{I,i}^*} - \ln \frac{\rho_{I,i}^*}{\rho_{I,i}} \right] \\
= & D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}^* \left(\lambda \frac{\rho_{S,j}}{\rho_{S,j}^*} + \lambda \ln \frac{\rho_{S,j}^*}{\rho_{S,j}} + \frac{\rho_{I,j}}{\rho_{I,j}^*} + \ln \frac{\rho_{I,j}^*}{\rho_{I,j}} \right) \\
& - D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}^* \left(\lambda \frac{\rho_{S,i}}{\rho_{S,i}^*} + \lambda \ln \frac{\rho_{S,i}^*}{\rho_{S,i}} + \frac{\rho_{I,i}}{\rho_{I,i}^*} + \ln \frac{\rho_{I,i}^*}{\rho_{I,i}} \right) \\
= & D_I \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{I,j}^* [G_j(\rho_{S,j}, \rho_{I,j}) - G_i(\rho_{S,i}, \rho_{I,i})], \tag{25}
\end{aligned}$$

where

$$G_i(\rho_{S,i}, \rho_{I,i}) = \lambda \frac{\rho_{S,i}}{\rho_{S,i}^*} + \lambda \ln \frac{\rho_{S,i}^*}{\rho_{S,i}} + \frac{\rho_{I,i}}{\rho_{I,i}^*} + \ln \frac{\rho_{I,i}^*}{\rho_{I,i}}. \tag{26}$$

Here we have used the fact that $1 - x + \ln x \leq 0$ for $x > 0$, and the equality holds if and only if $x = 1$.

Consider a weight matrix $W = (w_{ij})$ with entries $w_{ij} = D_I (A_{ji}/k_j) \rho_{I,j}^*$, and denote the corresponding weighted digraph as (G, W) . Let $l_i = \sum_{T \in \mathbb{T}_i} w(T) \geq 0$ be the same as that given in (A.1) in the Appendix associated with (\mathcal{L}, W) . Then, by (A.2), the following identity holds:

$$\sum_{i=1}^N l_i \sum_{j=1}^N D_I \frac{A_{ji}}{k_j} \rho_{I,j}^* [G_j(\rho_{S,j}, \rho_{I,j}) - G_i(\rho_{S,i}, \rho_{I,i})] = 0. \tag{27}$$

Set

$$V(\rho_{S,1}, \rho_{I,1}, \dots, \rho_{S,N}, \rho_{I,N}) = \sum_{i=1}^N l_i V_i(\rho_{S,i}, \rho_{I,i}). \tag{28}$$

Then, we obtain

$$\begin{aligned}
\frac{dV}{dt} \Big|_{(2.2)} &= \sum_{i=1}^N l_i \frac{dV_i}{dt} \Big|_{(2.2)} \\
&\leq \sum_{i=1}^N l_i \sum_{j=1}^N D_I \frac{A_{ji}}{k_j} \rho_{I,j}^* [G_j(\rho_{S,j}, \rho_{I,j}) - G_i(\rho_{S,i}, \rho_{I,i})] = 0, \tag{29}
\end{aligned}$$

for all $(\rho_{S,1}, \rho_{I,1}, \dots, \rho_{S,N}, \rho_{I,N}) \in \mathring{\Gamma}$. Since A is irreducible, we know that $l_i > 0$ for all i (see the Appendix), and thus $(dV/dt)|_{(2.2)} = 0$ implies that $\rho_{S,i} = \rho_{S,i}^*$ for all i . From the first equation of (3), we obtain

$$0 = b_i - \beta \rho_{S,i}^* \rho_{I,i} + \gamma \rho_{I,i} - D_S \rho_{S,i}^* + D_S \sum_{j=1}^N \frac{A_{ji}}{k_j} \rho_{S,j}^* - \mu \rho_{S,i}^*, \tag{30}$$

for $i = 1, \dots, N$, which implies that $\rho_{I,i} = \rho_{I,i}^*$. This means that the largest invariant set in $\{(dV/dt)|_{(2.2)} = 0\}$ is the singleton $\{E^*\}$. By LaSalle's invariance principle, E^* is globally asymptotically stable in $\mathring{\Gamma}$. \square

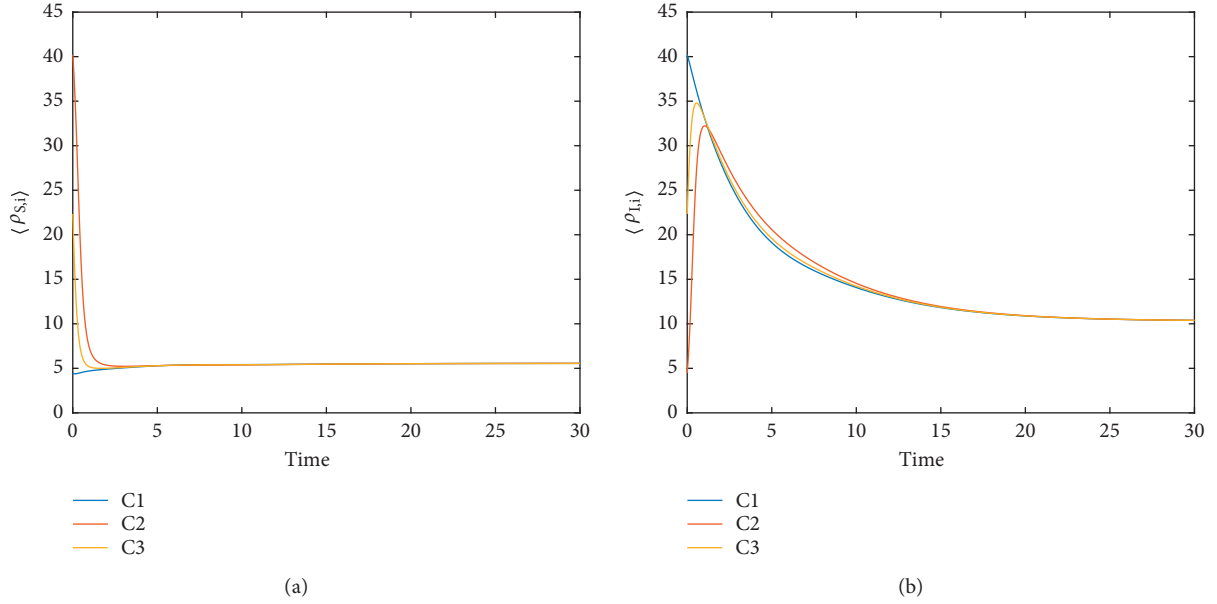


FIGURE 1: The average number of susceptible and infected individuals with different initial values.

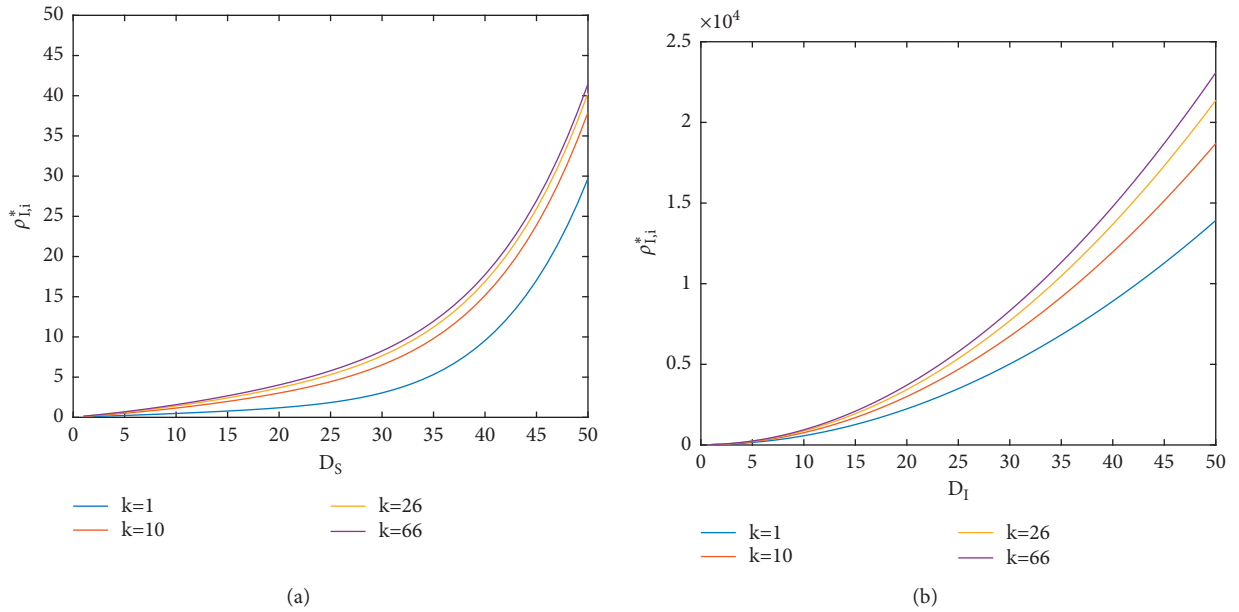


FIGURE 2: The impact of diffusion on the number of infected individuals at the endemic equilibria, and each line corresponds to the nodes with degrees k of 1, 10, 26, and 66, respectively (from bottom to top). (a) Relationship between D_S and $\rho_{I,i}^*$. (b) Relationship between D_I and $\rho_{I,i}^*$.

3. Numerical Simulations

The results in the previous section presented the stability of the disease-free and endemic equilibria. Through numerical simulations of infection dynamics, it is shown that the connectivity between patches has a large effect on disease dynamics. In this section, an Erdős–Rényi network with 100 nodes is established to study the influence of diffusion, and the connection probability is 0.4. All parameters are positive.

It should be noted that we simulated the numerical solution of the model equation, not Monte Carlo (MC) simulations of the metapopulation dynamics [14].

Figure 1 shows the average number of susceptible and infected individuals with different initial values. The parameters are used as $N = 100$, $b = 1$, $\mu = 0.00312$, $\beta = 0.125$, $D_S = 0.2$, $D_I = 0.2$, and $\gamma = 0.6$. We select three different initial conditions for simulation, and the three initial values were C1: the ratio of susceptible number to infected number

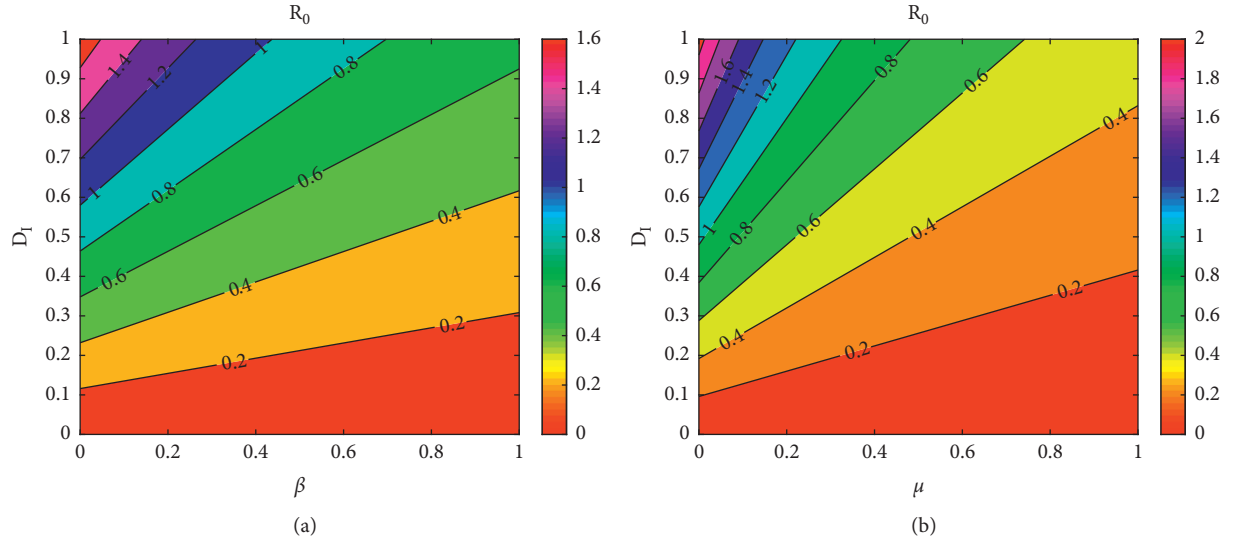


FIGURE 3: The impact of diffusion and other parameters on R_0 . (a) Relationship between β , D_I , and R_0 . (b) Relationship between μ , D_I , and R_0 .

is 1:9; C2: the ratio of susceptible number to infected number is 1:1; and C3: the ratio of susceptible number to infected number is 9:1. Susceptible and infected individuals obey Poisson distribution in nodes. It can be seen in Figure 1 that under different initial conditions, both susceptible individuals and infected individuals converge to a unique globally asymptotically stable endemic equilibrium.

In Figure 2, we consider the impact of diffusion on the number of infected individuals at the endemic equilibrium. For this purpose, we take $N = 100$, $b = 6$, $\mu = 0.00512$, $\beta = 0.325$, and $\gamma = 0.6$. In Figure 2(a), we fix $D_I = 1$. We observe that $\rho_{1,i}^*$ increases as D_S increases and for the same D_S , when i is larger, $\rho_{1,i}^*$ is also larger. In Figure 2(b), we take $D_S = 2$. It is observed that the same phenomena happen with respect to D_I and i with fixed D_I .

The relationship between β , D_I , and R_0 is simulated in Figure 3(a), and the relationship between μ , D_I , and R_0 is simulated in Figure 3(b). It can be seen from Figure 3 that the influence of D_I on R_0 is greater than that of β and μ , so the influence of the diffusion rates for the infected individuals on disease outbreak cannot be ignored.

4. Conclusion

In this paper, we formulated and analyzed an SIS model with demographics and diffusion on metapopulation networks. In this model, we solved the stability of the disease-free equilibrium and deduced the basic reproduction number. We concluded that when $R_0 < 1$, the disease-free equilibrium is globally asymptotically stable; otherwise, it is unstable. Next, we proved the existence and stability of endemic equilibria. The endemic equilibrium is globally asymptotically stable when $R_0 > 1$. Finally, we gave some numerical simulations to illustrate the main results and demonstrate the effects of connectivity and diffusion on the number of infected individuals at the endemic equilibrium. As the connectivity or diffusion is increasing,

the number of infected individuals is also increasing. This means that when an epidemic comes, the more people there are and the greater the degree is, the more likely the disease will break out. Therefore, we should pay attention to the prevention and control of diseases and make plans in advance.

The research of this paper can be extended to the following aspects: in this paper, the Erdős-Rényi network is used to simulate, and the models based on other networks can be analyzed; secondly, the infectious disease model with heterogeneous diffusion rate can also be analyzed. It is necessary to study the mathematical model under a specific network in the future, so as to make the model more realistic.

Appendix

A. Combinatorial Identity

Let (\mathcal{G}, W) be a weighted digraph with $N (> 2)$ vertices, where $W = (w_{ij})$ is the weight matrix. A weight $w_{ij} > 0$ if the directed arc (j, i) from vertex j to vertex i exists; otherwise, $w_{ij} = 0$. Let \mathbb{T}_i be the set of all spanning trees of (\mathcal{G}, W) rooted at vertex i . For $T \in \mathbb{T}_i$, the weight of T , denoted by $w(T)$, is the product of weights on all arcs of T . Let

$$l_i = \sum_{T \in \mathbb{T}_i} w(T), \quad i = 1, 2, \dots, N. \quad (\text{A.1})$$

Then, $l_i \geq 0$ and

$$\sum_{i,j=1}^N l_i w_{ij} G_i(x_i) = \sum_{i,j=1}^N l_i w_{ij} G_j(x_j), \quad (\text{A.2})$$

for any family of functions $\{G_i(x_i)\}_{i=1}^N$. If $W = (w_{ij})$ is irreducible, then $l_i > 0$ for $i = 1, 2, \dots, N$. We refer readers to [22] for the proof of (A.1).

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 are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (nos. 12071445 and 12001501), Fund for Shanxi (1331KIRT), Shanxi Natural Science Foundation (nos. 201801D221033 and 201901D211216), Graduate Innovation Project in Shanxi Province (2020BY104), and the Outstanding Youth Fund of North University of China.

References

- [1] N. Masuda, "Effects of diffusion rates on epidemic spreads in metapopulation networks," *New Journal of Physics*, vol. 12, 2010 093009.
- [2] L. Longini and I. L. Jr, "A mathematical model for the global spread of influenza," *Mathematical Biosciences*, vol. 75, no. 1, pp. 3–22, 1985.
- [3] I. Hanski and O. Ovaskainen, "The metapopulation capacity of a fragmented landscape," *Nature*, vol. 404, no. 6779, pp. 755–758, 2000.
- [4] S. Bonaccorsi, S. Ottaviano, D. Mugnolo, and F. D. Pellegrini, "Epidemic outbreaks in networks with equitable or almost-equitable partitions," *SIAM Journal on Applied Mathematics*, vol. 75, no. 6, pp. 2421–2443, 2015.
- [5] 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, 2008.
- [6] C. Poletto, S. Meloni, V. Colizza, Y. Moreno, and A. Vespignani, "Host mobility drives pathogen competition in spatially structured populations," *PLoS Computational Biology*, vol. 9, no. 8, Article ID 1003169, 2013.
- [7] K. M. A. Ksbir and J. Tanimoto, "Evolutionary vaccination game approach in metapopulation migration model with information spreading on different graphs," *Chaos, Solitons & Fractals*, vol. 120, pp. 41–55, 2019.
- [8] W. Wang, X. Gao, Y. Cai, H. Shi, and S. Fu, "Turing patterns in a diffusive epidemic model with saturated infection force," *Journal of the Franklin Institute*, vol. 355, no. 15, pp. 7226–7245, 2018.
- [9] G. Luca, K. V. Kerckhove, P. Coletti et al., "The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium," *BMC Infectious Diseases*, vol. 18, no. 1, 29 pages, 2018.
- [10] I. Hanski and O. E. Gaggiotti, "Ecology, genetics and evolution of metapopulations," *Journal of the Torrey Botanical Society*, vol. 131, no. 4, 2004.
- [11] M. J. Keeling and P. Rohani, "Estimating spatial coupling in epidemiological systems: a mechanistic approach," *Ecology Letters*, vol. 5, no. 1, pp. 20–29, 2002.
- [12] A. Ewing, E. C. Lee, C. Viboud, and S. Bansal, "Contact, travel, and transmission: the impact of winter holidays on influenza dynamics in the United States," *Journal of Infectious Diseases*, vol. 215, no. 5, pp. 732–739, 2017.
- [13] J. Saldaña, "Continuous-time formulation of reaction-diffusion processes on heterogeneous metapopulations," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 78, Article ID 012902, 2008.
- [14] D. Juher, J. Ripoll, and J. Saldaña, "Analysis and Monte Carlo simulations of a model for the spread of infectious diseases in heterogeneous metapopulations," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 80, Article ID 041920, 2009.
- [15] J. Wang, M. Liu, and Y. Li, "Analysis of epidemic models with demographics in metapopulation networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 7, pp. 1621–1630, 2013.
- [16] L. Perko, *Differential Equations and Dynamical Systems*, Springer-Verlag, Berlin, Germany, 2000.
- [17] M. Y. Li and Z. Shuai, "Global stability of an epidemic model in a patchy environment," *Canadian Applied Mathematics Quarterly*, vol. 17, no. 1, pp. 175–187, 2009.
- [18] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, Cambridge, MA, USA, 1979.
- [19] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Mathematical Biosciences*, vol. 180, no. 1–2, pp. 29–48, 2002.
- [20] M. Y. Li, J. R. Graef, L. Wang, and J. Karsai, "Global dynamics of a SEIR model with varying total population size," *Mathematical Biosciences*, vol. 160, no. 2, pp. 191–213, 1999.
- [21] N. P. Bhatia and G. P. Szego, *Dynamical Systems: Stability Theory and Applications*, Springer-Verlag, Berlin, Germany, 1967.
- [22] F. Brauer, P. van den Driessche, and L. Wang, "Oscillations in a patchy environment disease model," *Mathematical Biosciences*, vol. 215, no. 1, pp. 1–10, 2008.
- [23] H. Guo, M. Y. Li, and Z. Shuai, "Global stability of the endemic equilibrium of multigroup SIR epidemic models," *Canadian Applied Mathematics Quarterly*, vol. 14, pp. 259–284, 2006.
- [24] H. Guo, M. Y. Li, and Z. Shuai, "A graph-theoretic approach to the method of global Lyapunov functions," *Proceedings of the American Mathematical Society*, vol. 136, no. 8, pp. 2793–2802, 2008.
- [25] M. Y. Li and Z. Shuai, "Global-stability problem for coupled systems of differential equations on networks," *Journal of Differential Equations*, vol. 248, no. 1, pp. 1–20, 2010.

Research Article

A Heterogeneous Branching Process with Immigration Modeling for COVID-19 Spreading in Local Communities in China

Lin Zhang ¹, Haochen Wang,² Zhongyang Liu,³ Xiao Fan Liu ⁴, Xin Feng ⁵ and Ye Wu^{6,7}

¹School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

²International School, Beijing University of Posts and Telecommunications, Beijing 100876, China

³School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China

⁴Department of Media and Communication, City University of Hong Kong, Kowloon, Hong Kong 999077, China

⁵School of Economics and Management, Yanshan University, Qinhuangdao 066004, China

⁶School of Information Engineering, Putian University, Putian 351100, China

⁷School of Journalism and Communication, Beijing Normal University, Beijing 100875, China

Correspondence should be addressed to Xiao Fan Liu; xf.liu@cityu.edu.hk and Xin Feng; 149987543@qq.com

Received 30 October 2020; Revised 24 March 2021; Accepted 7 May 2021; Published 22 May 2021

Academic Editor: Peter Giesl

Copyright © 2021 Lin 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 COVID-19 pandemic spread catastrophically over the world since the spring of 2020. In this paper, a heterogeneous branching process with immigration is established to quantify the human-to-human transmission of COVID-19 in local communities, based on the temporal and structural transmission patterns extracted from public case disclosures by four provincial Health Commissions in China. With proper parameter settings, our branching model matches the actual transmission chains satisfactorily and, therefore, sheds light on the underlying COVID-19 spreading mechanism. Moreover, based on our branching model, the efficacy of home quarantine and social distancing are explored, providing a reference for the effective prevention of COVID-19 worldwide.

1. Introduction

The COVID-19 spread alarmingly fast in Wuhan in late January 2020 before the city's lockdown starting Jan. 23. Based on the public reports on the number of confirmed cases, the prevalence of COVID-19 outside Hubei Province came to a controllable size in late February. The Wuhan lockdown eventually ended on Apr. 8, 76 days since its commencement, after the confirmation of COVID-19 under control in China.

An extensive amount of research has been conducted to understand the spreading features of COVID-19. There are two primary directions. One is the clinic feature, focusing on the virus itself [1, 2], such as the estimation of the basic reproduction number R_0 [3–5], the effective reproduction number R_t [6], and the basic statistical results obtained from the confirmed cases, such as the incubation period, the serial time, and the secondary attack rate [7–9]. The other

direction is the spreading dynamics, which is studied mainly through mathematical models, such as SEIR [10–14] and branching model [15, 16]. The effect of lockdown of Wuhan [17, 18] and different levels of isolations are also considered based on generalized or specialized SEIR models [15].

In this paper, public reports of line-list confirmed cases in Anhui, Henan, Jiangsu, and Zhejiang provinces from Jan. 21 to Feb. 19, 2020, covering 30 days, were collected and analyzed. Due to effective isolation policies, such as suggesting people to stay at home, wearing masks, washing hands, and tracing close contacts, the epidemic got under control within about one month in the above four provinces, which is approximately two or three generations according to the serial time. Short transmission chains are not appropriate to be modeled by the SEIR model, which usually simulates transmissions using multiple iterations. Therefore, instead of SEIR, we propose a branching process to model the spreading of COVID-19 in well-prevented regions in

China. Based on the statistical results extracted from our data, two influential factors for the propagation of COVID-19 are considered: the migration from outside a particular community and the efficacy of containment within the local communities.

In fact, SEIR models and branching models are both superior candidates for modeling classical epidemic spreading. Under certain conditions, such as when the total population is large enough, the two models are equivalent in modeling general epidemic dynamics; see [19] for theoretical support. In SEIR models, transmissions may happen in multiple iterations and result in longer transmission chains than the actual situation. In contrast, the branching process is more flexible in modeling the efficacy by the cutoff of the transmission chain in model assumption. On the contrary, the branching model is also flexible in modeling the confirmed cases of COVID-19 with different sources of contact, that is, imported or local, which are modeled by immigration and branching parts, respectively. In SEIR models, it is difficult to distinguish the sources of contact. The above comparison is listed in Table 1.

To sum up, instead of the well-known SEIR model, a heterogeneous branching process with immigration is established to explore the diffusion of COVID-19 in well-prevented local communities in China. In our branching model, heterogeneity is caused by the distribution of serial time, immigration is the confirmed cases coming to a certain local community from outside, and the secondary cases infected by the imported infectors are modeled as their offsprings with the specific branching mechanism. Further transmissions are modeled as further offsprings with similar rules. All parameters in the model are extracted and approximated from real data. The feasibility of this approach is verified by back analysis of choosing proper parameters which represent isolation strength and social distance. It turns out that our model matches the real data very well. The efficacy of the containment measures is also simulated with our branching model. Our findings reveal the spreading mechanism of the COVID-19 from an individual to the population level in well-prevented local communities. The effectiveness of isolation measures in local communities obtained in our work can shed light on preventing the global pandemic spreading of COVID-19.

An outline of this paper is as follows. The data description is given in Section 2. The branching model is built in Section 3, with parameters obtained by statistical analysis from the real data. The validation of our branching model and the impact analysis of the isolation parameters in our model are explored in Section 4. Conclusions and discussions are given in Section 5.

2. Data Description

The data in this paper are extracted from the reports of confirmed cases collected in Anhui (totally 887 cases), Henan (totally 1279 cases), Jiangsu (totally 577 cases), and Zhejiang (totally 1137 cases) from Jan. 21 to Feb. 19, 2020. The locations of these four provinces, as well as Hubei, are illustrated in Figure 1. The color refers to the number of

confirmed cases we collected in each region till Feb. 19, 2020. A typically reported item is as follows.

“Patient ID: Huainan-25. The patient Huainan-25 is a 59 year-old woman who is the wife of the Huainan-26 patient. On Feb. 12, she developed fever, muscle soreness, and other symptoms. On Feb. 14, she went to the hospital for treatment and stayed at the hospital for observation. On Feb. 15, her nucleic acid test was deemed positive, and doctors diagnosed her as a suspected patient. Two days later, she was confirmed. Doctors have traced back three close contacts, all of whom have been quarantined for medical observation. During the Chinese New Year’s holiday, she had close contact with her daughter, son-in-law, and granddaughter. Her son-in-law, an asymptomatic patient with a history of suspicious exposure in Hefei, stayed at a designated hospital for observation. Doctors have traced back his 46 close contacts, all of whom have been quarantined for medical observation.”

For the patient with ID “Huainan-25,” Huainan refers to the city where the patient lives, and 25 means that she is the 25th confirmed case in Huainan. The cases we selected are the ones with partial or full of the following information: (1) date of confirmation, (2) whether or not be an imported case (that is, infected outside the local community or not), (3) date of his/her infector’s confirmation, and (4) relationship between a primary case (infector) and a secondary case (infectee). After extracting the necessary information we need, sample sizes for (1) and (2) are 831 for Anhui, 967 for Henan, 299 for Jiangsu, and 1,051 for Zhejiang, respectively. For (3) and (4), 411 cases (with Anhui 234 and Henan 177) are obtained.

Based on the actual data, statistical results concerning the key features during the spreading are illustrated in Section 3, including (1) the imported and local new cases evolving with time, (2) the main relationships between infector and infectee, and (3) the serial interval, that is, the time interval of confirmation times between each pair of infector and infectee.

3. Model Description: Heterogeneous Branching Process with Immigration

Naturally, a strict isolation policy is urgently needed for highly infectious diseases without pharmaceutical measures to prevent its epidemic effectively. The detailed reports of confirmed cases provide necessary information to understand the mechanism of COVID-19 transmission. The tracing back and isolation of close contacts efficiently cut off the transmission chain such that the imported cases to a specific region could only transmit the virus for few more generations. The serial interval and the incubation period are two of the key factors for prevention policymaking, from which the suggested length of isolation is commonly set as at least 14 days.

A heterogeneous branching process with immigration is established considering three ingredients for modeling, which are (1) the temporal pattern of serial time, (2) the structural pattern of transmission considering containment measures, and (3) the import of confirmed cases which

TABLE 1: The comparison of SEIR and branching in modeling COVID-19 in well-prevented regions.

| Model | SEIR | Branching |
|------------|---|--|
| Merit | Simple and easy for modeling the epidemic spreading with mean field approach under well mixture assumption | (i) Flexible for modeling different sources of contact (ii) Flexible for specific transmission rules during the spreading |
| Limitation | (i) Iteration procedure results in longer transmission chains (ii) It is difficult for modeling the agents with different sources of contact | The simulating procedure is more complex for tracing every transmission tree |



FIGURE 1: The location of Hubei, Anhui, Henan, Jiangsu, and Zhejiang. The color refers to the number of confirmed cases we collected in each region till Feb. 19, 2020. That is, Anhui, 887 cases, Henan 1279 cases, Jiangsu 577 cases, and Zhejiang 1137 cases.

begins the prevalence of COVID-19 in local communities. The framework of our branching model is given in Section 3.1. The values of parameters and distributions of random variables in our branching model are extracted from the real data in Sections 3.2, 3.3, and 3.4. The validation of our model and simulation results for different isolation levels and social distances are given in Section 4.

3.1. The Framework of the Branching Model. Heterogeneous branching processes with immigration are well suited to describe the temporal evolution of populations in which individuals appear randomly over time in accordance with two distinct mechanisms. One mechanism, called immigration, is the influx of new individuals in the population of which they are not natives. The other mechanism, referred to as branching, is how individuals of the population generate new offspring. In this paper, we consider a heterogeneous branching process with immigration, in which

- (i) the branching mechanism is used to model the spreading of the virus in local communities with heterogeneity caused by serial time

- (ii) the immigration is a time-dependent Poissonian process, modeling the imported cases coming from the outside of a certain region

In the following, immigration, offspring distribution, and serial time are discussed in detail with values or distributions obtained from statistical analysis of the real data described in Section 2.

3.2. Immigration. The imported cases with contact history from outside of a local region are described as immigration. In our branching model, the immigration process is modeled by a time-dependent Poissonian process with a varying rate $r(t)$. That is, the number of immigration arrived on the t th day, denoted as $I(t)$ for $t = 1, 2, \dots$, possesses the following distribution law:

$$P(I(t) = k) = \frac{r(t)^k}{k!} e^{-r(t)}, \quad k = 0, 1, \dots \quad (1)$$

From the line-list reports, the number of imported cases and the number of local cases changing with time are

obtained. Figure 2 illustrates the data results of the four provinces we considered. The red and black curves refer to the imported and local cases, respectively. The immigration process established in our model is extracted from the imported sequences of the four provinces, referring to the red curves in Figure 2.

It is apparent that the first spreading in local communities is due to the import of confirmed cases from outside of the considered local community. In the beginning, the imported cases are more than the local ones. Then, several days later, local cases began to increase. Whether or not an outbreak will happen depends on the prevention policy of local communities as long as the import path is completely cut off at an early stage.

3.3. Offspring. The number of potential secondary cases produced by each infective individual is called the offspring in our branching model, which comprises two parts considering the place where infection of COVID-19 happens.

One is within a family, drawn from a binomial distribution $\text{Bin}(N-1, p)$, where N is a random variable representing the number of family members and p is the probability of getting infected within a family by the first infected member. It is notable that, for $N=1$, there is no other family member to be infected. Moreover, high transmission of COVID-19 results in a large value of p . In our model, the transmission within a family happens with probability $p=0.9$ according to the statistical result that about 90% infection happened between family members.

The relationship between each pair of infector and infectee is counted. Due to home quarantine and high transmissibility of the COVID-19, the family members of the imported infectious ones are at super high risk of being infected. The top three relationships between an infector and an infectee are between couples, from parent to child, and from child to parent. The number and ratio of cases for the three relationships are shown in Table 2.

For the number of family members N , the reference distribution comes from the Chinese statistical yearbook of 2020. The distribution is illustrated in Table 3.

The other part of the offspring happens out of their homes. Assume the probability of leaving home is $\alpha \in (0, 1)$. Moreover, the number of potential infectees outside homes follows Poissonian distribution $\text{Poi}(\lambda)$, $\lambda > 0$. In other words, α represents the strength of home quarantine, and λ measures the effect of social distance. Smaller α and/or λ means more strict containment of COVID-19 in certain regions.

The final assumption comes from the isolation and tracing back policy. The secondary infectors' behavior is slightly adjusted. Firstly, the family infectees would not transmit the virus to family members since they are all treated as the offspring of the first infected family member who imported the virus. Secondly, the family infectees may have secondary out-of-home infectees, but the probability changes from α to α^2 . In fact, the decaying pattern of the probability of going out of home is set as exponential due to the cumulative awareness of isolation. Therefore, the

probability of leaving home for the second generation is set as α^2 instead of linear relation or others between generations. Thirdly, for the social infectees of the imported cases initially, they can transmit the virus to their family members, and they may also have their secondary out-of-home infectees. The probability also changes from α to α^2 . Finally, no more transmission would happen after two generations due to the strict contact tracing measures taken at local communities.

3.4. Heterogeneity. The heterogeneity in our branching model comes from the serial time, denoted by T , which is the time interval between the onset times of a newly infectee and its infector. The serial interval distribution extracted from the data is illustrated in Figure 3.

The empirical distribution of T in Figure 3 is obtained with the 80.05% positive serial intervals. It is notable that there are also negative and zero serial intervals. The ratios of negative and zero serial intervals are 5.35% and 14.60% in our sample, shown in the bar plot in Figure 3. Together with the empirical distribution of the positive serial interval, two known distributions, which are the Gamma distribution with mean 4.43 and variance 10.23 and the Weibull distribution with mean 4.45 and variance 10.31, are utilized to fit the empirical distribution. The fitting distributions are drawn in Figure 3 as references. Notably, a translational Weibull distribution is utilized in Ref. [6] for the serial time with different datasets, which is consistent with our result. The numerical result of our empirical distribution is listed in Table 4 for reference, which is used in our simulation.

To sum up, the parameters or variables of our branching model, together with their descriptions and values or distributions for further investigation and simulation, are listed in Table 5. Notably, the tested parameters are α and λ , representing the isolation level and social distance. Other parameters or variables involved in our model are kept fixed during the simulation.

4. Simulation Results

Firstly, we show the good match of our model to the real data in Section 4.1. Then, the efficiency of staying at home with parameter α and keeping social distance with parameter λ is provided in Section 4.2. In the following simulation, the distributions of family members N and the serial interval T are listed in Table 5, and the transmission probability between family members p is fixed as 0.9.

4.1. Fitting Real Spreading Processes. To match the real data, the imported data series is borrowed as the immigration of the branching process in each province, which are the red curves in Figure 2. In order to test the validation of our branching structure, the simulation results should match the black curves in Figure 2. Therefore, fine values for the tested parameters, i.e., the probability of going out of home α and the mean of secondary cases due to social activities λ should be set carefully. The best fit values for α and λ are obtained by minimizing the mean absolute error (MAE) between the real

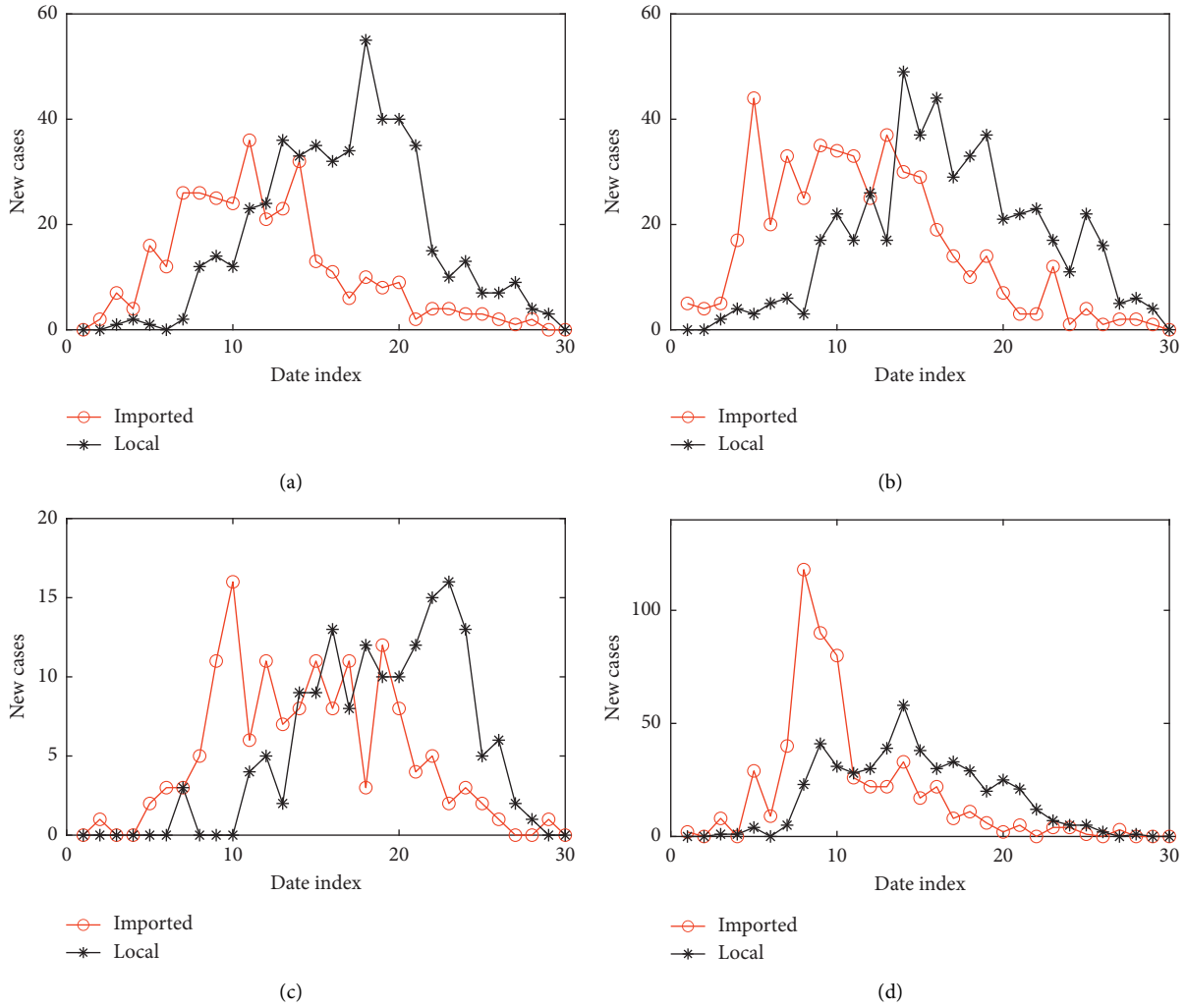


FIGURE 2: The number of imported and local confirmed cases changing with time for (a) Anhui Province (831 cases), (b) Henan Province (967 cases), (c) Jiangsu Province (299 cases), and (d) Zhejiang Province (1051 cases). The red and black curves are the imported and the local cases, respectively.

TABLE 2: The top three infector-infectee relationships within family members based on 411 confirmed cases.

| Relationship | Couples | Parent-child | Child-parent |
|--------------|---------|--------------|--------------|
| Counts | 97 | 47 | 43 |
| Probability | 0.52 | 0.25 | 0.23 |

TABLE 3: The reference distribution of the number of family members N from Chinese statistical yearbook of 2020.

| N | 1 | 2 | 3 | 4 | 5 | ≥ 6 |
|-------------|------|------|------|------|------|----------|
| Probability | 0.18 | 0.30 | 0.22 | 0.16 | 0.08 | 0.06 |

local series and the simulated ones. The simulation series is obtained by averaging the 50 experiment trials.

Figure 4 shows the simulation results. As shown in Figure 4, the simulated local confirmed case series and the real ones match well for all the four provinces, which gives good validation of our model, proving that the branching structure built in this paper is adequate for modeling the

spreading of COVID-19 in well-prevented local communities.

4.2. *Simulation Results.* In this section, experiments are conducted to investigate the combined effect of staying at home and keeping social distance. For this purpose, the

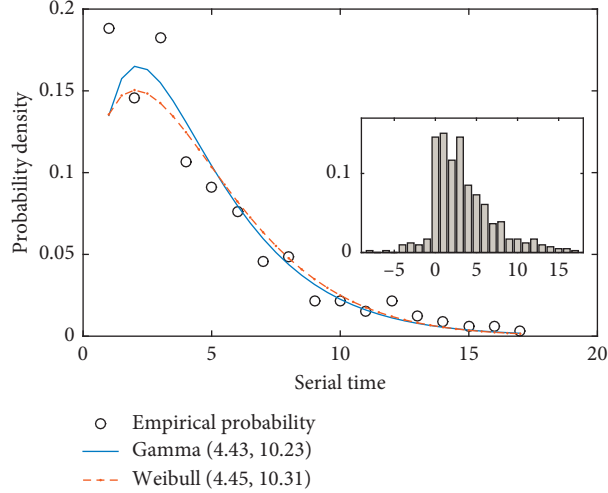


FIGURE 3: The empirical distribution of the positive serial time with 329 confirmed cases. The reference fitted distributions are Gamma distribution with mean 4.43 and variance 10.23 (the blue line) and the Weibull distribution with mean 4.45 and variance 10.31 (the red dash-dotted line). The inserted is the histogram of all serial times for 411 confirmed cases, containing nonpositive ones.

TABLE 4: Empirical distribution law of the positive serial interval T .

| T | Frequency |
|-----|-----------|
| 1 | 0.1884 |
| 2 | 0.1459 |
| 3 | 0.1824 |
| 4 | 0.1064 |
| 5 | 0.0912 |
| 6 | 0.0760 |
| 7 | 0.0456 |
| 8 | 0.0486 |
| 9 | 0.0213 |
| 10 | 0.0213 |
| 11 | 0.0152 |
| 12 | 0.0213 |
| 13 | 0.0122 |
| 14 | 0.0091 |
| 15 | 0.0061 |
| 16 | 0.0061 |
| 17 | 0.0030 |

immigration rate $r(t)$ is set as the average of the four series from the four provinces with a moving average of order 5, which is illustrated as the red curve with circles in Figure 5(b).

First of all, the spatial stratified heterogeneity (SSH) among provinces is measured. The program calculates a so-called q -statistic to test the significance of differences among provinces. The value of q is a ratio ranging from 0 to 1, where 0 means no association between the number of cases and province, while 1 means that they are perfectly associated. The q -statistic can be calculated with the following equation [20]:

$$q = 1 = \frac{SSW}{SST}, \quad (2)$$

where $SSW = \sum_{h=1}^L N_h \sigma_h^2$, $SST = N\sigma^2$, N and σ^2 are the numbers of units and the variance in the study area which is composed of L strata, respectively, N_h is the number of units,

and σ_h is the variance in stratum h . Large value of q means larger spatial heterogeneity in the study area. The significance value p can be transformed so that it can satisfy the noncentral F -distribution:

$$F = \frac{N-L}{L-1} \frac{q}{1-q} \sim F(L-1, N-L; \lambda), \quad (3)$$

with

$$\lambda = \frac{1}{\sigma^2} \left(\sum_{h=1}^L \bar{Y}_h^2 - \frac{1}{N} \left(\sum_{h=1}^L \sqrt{N_h} \bar{Y}_h \right)^2 \right), \quad (4)$$

where λ and \bar{Y}_h are the noncentral parameter and mean value in stratum h , respectively. Then, the q -statistic and the corresponding p value in Table 6 can be applied to testify that whether the concerned cases have significant differences of variances in different strata. As the p values in Table 6 and curves in Figure 5, the SSH for all cases are significant at the level slightly above 0.05. The only non-significant one is the new cases of immigrant. As can be seen from Figure 5, the extreme fluctuation of Zhejiang Province is the main reason that leads to this non-significance. To sum up, the SSH for the time series we considered is significant. Despite the SSH, our branching model can fit different cases quite well with different parameters. In the following simulation for the isolation parameter α and social distance parameter λ , the immigrant rate $r(t)$ is fixed as the reference, which is chosen as the moving average of order 5 of the mean immigrant series of the four provinces, just the red curve in Figure 5(b).

In the following, we conduct the simulation for our parameters. The values for isolation parameter α and social distance parameter λ for simulation are selected as $\alpha = 0.1, 0.4, \text{ and } 0.7$ and $\lambda = 0.4, 1.4, \text{ and } 4.8$. The simulation results will be illustrated in Figures 6 and 7, with 9 subfigures for the nine combination of parameters α and λ . Figure 6 shows the evolution of local cases (the black curves)

TABLE 5: Parameters or variables' descriptions with values or distributions for our model.

| Parameter | Value | Description |
|-----------|----------------------------|---|
| $r(t)$ | As in Figure 4 | Rate for Poisson immigration process $\{I(t)\}$ |
| N | As in Table 3 | The distribution of family members |
| p | 0.9 | Transmission probability within family members |
| α | 0.1, 0.4, and 0.7 | Probability of social activities out of home |
| λ | 0.4, 1.4, and 4.8 | Mean of Poisson-distributed infectees due to social contact |
| T | As in Figure 3 and Table 4 | Serial interval between confirmations |

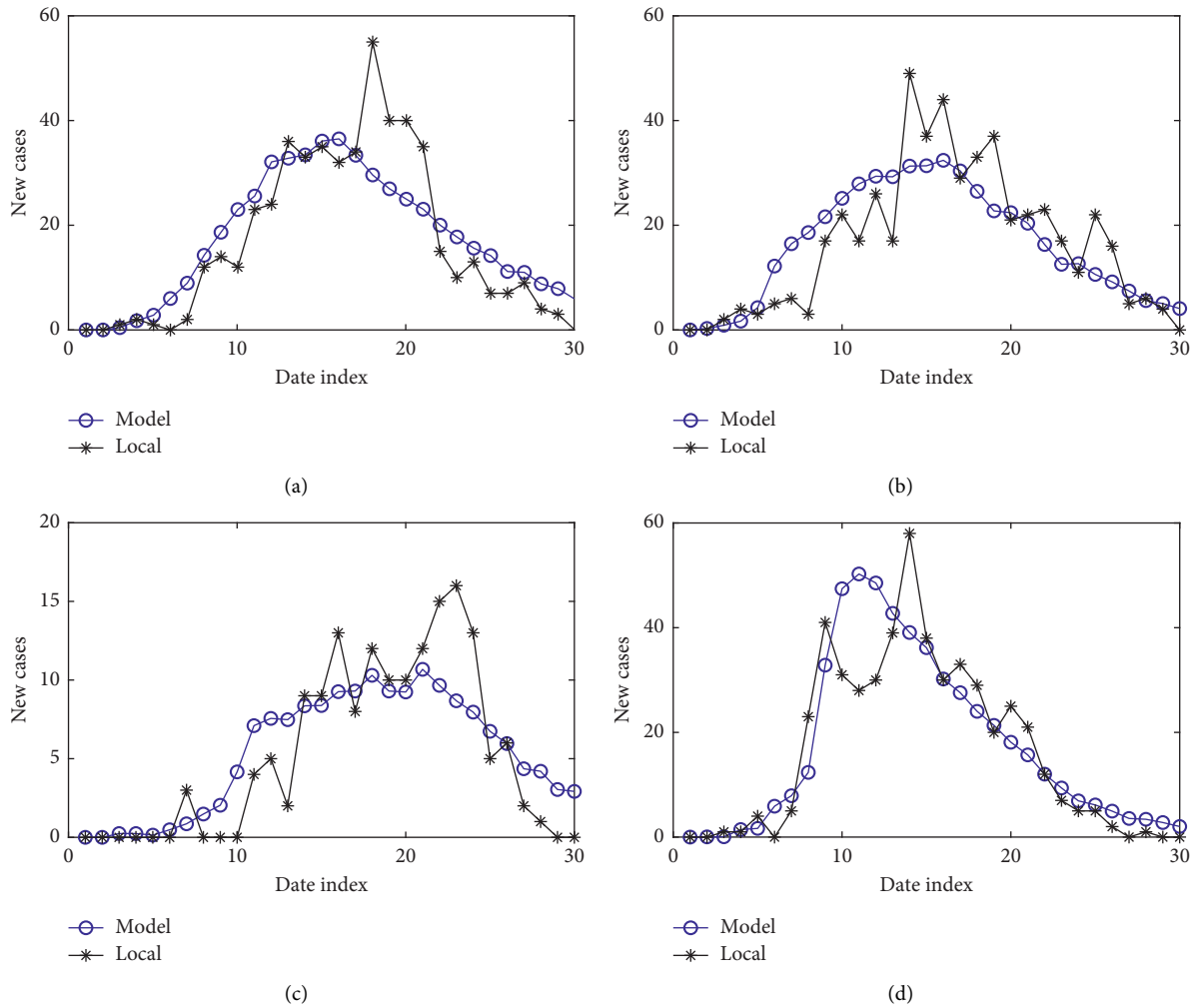


FIGURE 4: The simulating results and the real time series. The parameters for the best fit of the local series in each province are provided on top of each figure. The blue curves are obtained from our model with the average of 50 trials. The black curves are the original time series. (a) Anhui ($\alpha = 0.8$ and $\lambda = 0.4$). (b) Henan ($\alpha = 0.1$ and $\lambda = 1.0$). (c) Jiangsu ($\alpha = 0.7$ and $\lambda = 0.2$). (d) Zhejiang ($\alpha = 0.1$ and $\lambda = 0.1$).

changing with time, with the same immigration curve (the red ones) as the reference. Figure 7 gives the detailed components of local infectees by Home, Social, and Secondary. Home and Social refer to the infected cases of the imported cases taking place at home and out of home, respectively. Secondary refers to the infectees caused by Home and Social. The red, blue, and black curves in Figure 7 are the local infectees of Home, Social, and Secondary, respectively. Based on our assumption, the branching model only evolves two generations due to the contact tracing policy. In the

following, detailed results with isolation parameter α and social distance parameter λ are given.

Firstly, either strict isolation or keeping a strict social distance is effective for preventing the spreading. The effect of strict isolation is obtained from in Figures 6(a)–6(c), in which $\alpha = 0.1$. Obviously, the confirmed local cases (black curves) grow as λ increases but still within a controllable size. In Figures 7(a)–7(c), the numbers of Home (red curves) keep stable, while the Social (black curves) and Secondary (blue curves) infectees increase slightly as the gathering together

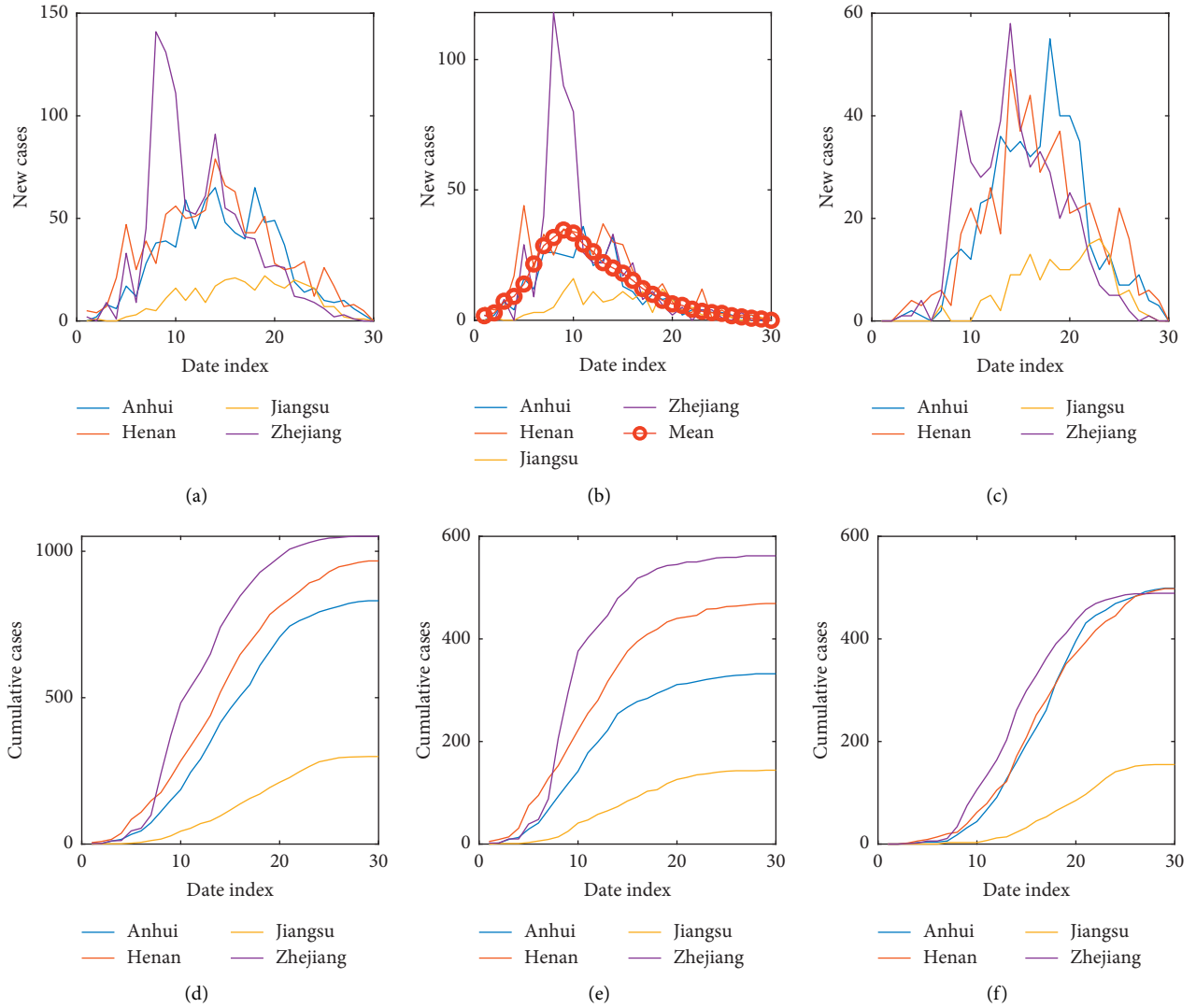


FIGURE 5: The time series for new cases and cumulative cases for different provinces. Legends for provinces are shown in subfigure (b). The total, immigrant, and local cases are considered, respectively. The rate $r(t)$ of immigrant $I(t)$ for simulations, which is the moving average of order 5 of the mean immigrant series of the four provinces, is as the red curve in (b). (a) Total. (b) Immigrant. (c) Local. (d) Total. (e) Immigrant. (f) Local.

TABLE 6: The q -statistic and corresponding p value for the SSH test.

| | | Total | Immigrant | Local |
|------------------|----------------|----------|-----------------------|---------|
| New cases | q -statistic | 0.1350 | 0.0896 | 0.1191 |
| | p value | 0.0237 | 0.0647 | 0.0406 |
| Cumulative cases | q -statistic | 0.2677 | 0.3630 | 0.1816 |
| | p value | 0.000131 | 8.16×10^{-7} | 0.00407 |

parameter λ increases. Therefore, the most effective measure for preventing the spreading is staying at home for about two weeks.

Secondly, when it is necessary to leave home, keeping a social distance is the second line of defense. Since it is difficult to stay at home for a couple of weeks without going out, well prevention is crucial to avoid being infected. The effectiveness of social distance can be obtained from Figures 6(a), 6(d), and 6(g), in which $\lambda = 0.4$. The local cases

increase as α increases but still within a controllable size. In Figures 7(a), 7(d), and 7(g), the numbers of Social (black curve) and Secondary (red curve) increase slightly as the leaving home probability α increases. However, as long as the social distance is far enough, isolation can be mitigated.

Finally, if isolation fails, for illustration, people have high demands of going out of their home, it is crucial to keep social distance, or the disaster result of gathering together will merge. As shown in Figures 6(g)–6(i) with $\alpha = 0.7$, the

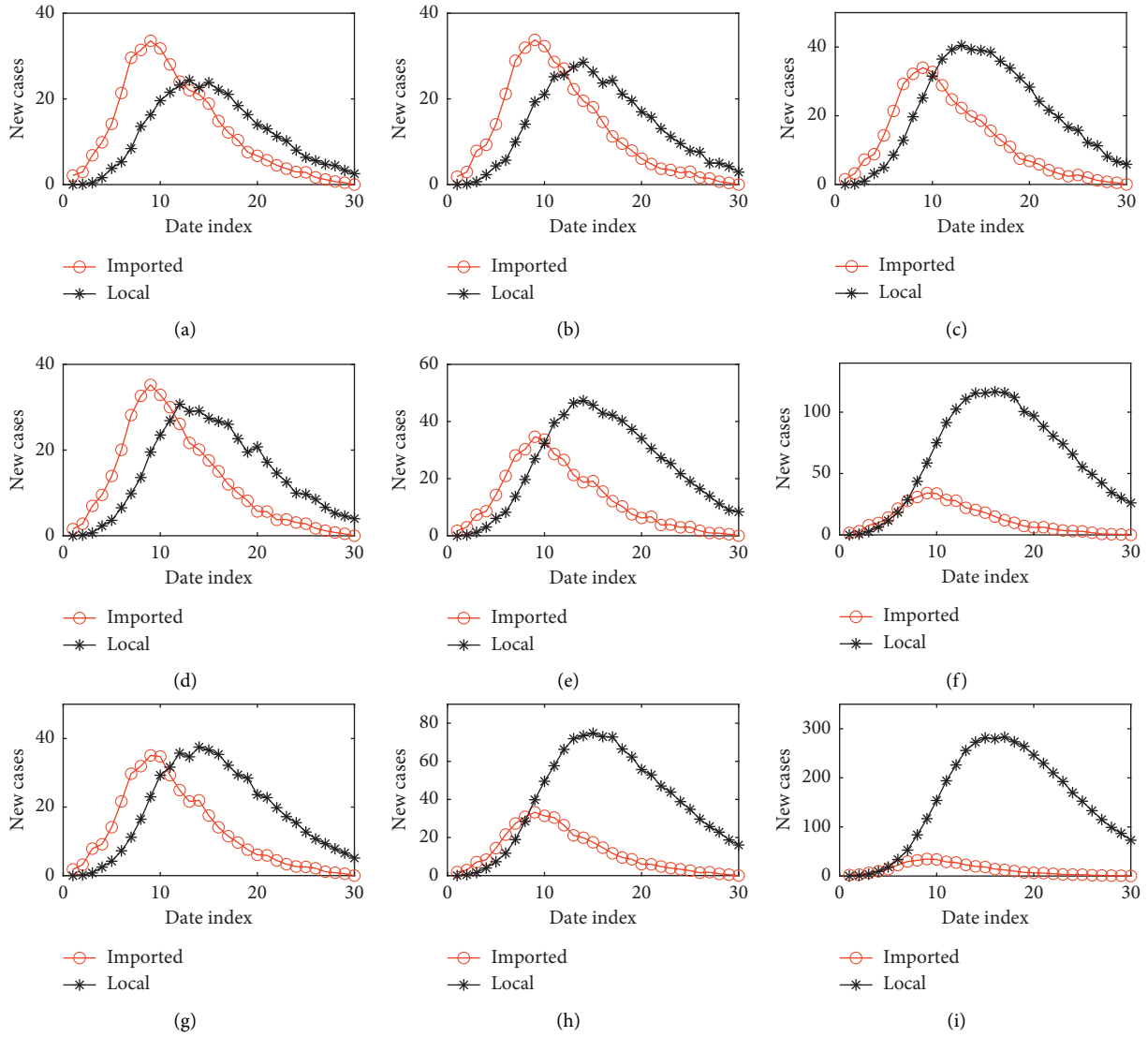


FIGURE 6: The simulation results of the age-dependent branching process with immigration. The red and black curves are for the imported and local cases, respectively. The red curves in all the subfigures are simulated with the same rate $r(t)$, as shown in Figure 5. The black curves are the number of local cases with different combinations of the values for the testing parameters as $\alpha = 0.1, 0.4$, and 0.7 and $\lambda = 0.4, 1.4$, and 4.8 . (a) $\alpha = 0.1$ and $\lambda = 0.4$. (b) $\alpha = 0.1$ and $\lambda = 1.4$. (c) $\alpha = 0.1$ and $\lambda = 4.8$. (d) $\alpha = 0.4$ and $\lambda = 0.4$. (e) $\alpha = 0.4$ and $\lambda = 1.4$. (f) $\alpha = 0.4$ and $\lambda = 4.8$. (g) $\alpha = 0.7$ and $\lambda = 0.4$. (h) $\alpha = 0.7$ and $\lambda = 1.4$. (i) $\alpha = 0.7$ and $\lambda = 4.8$.

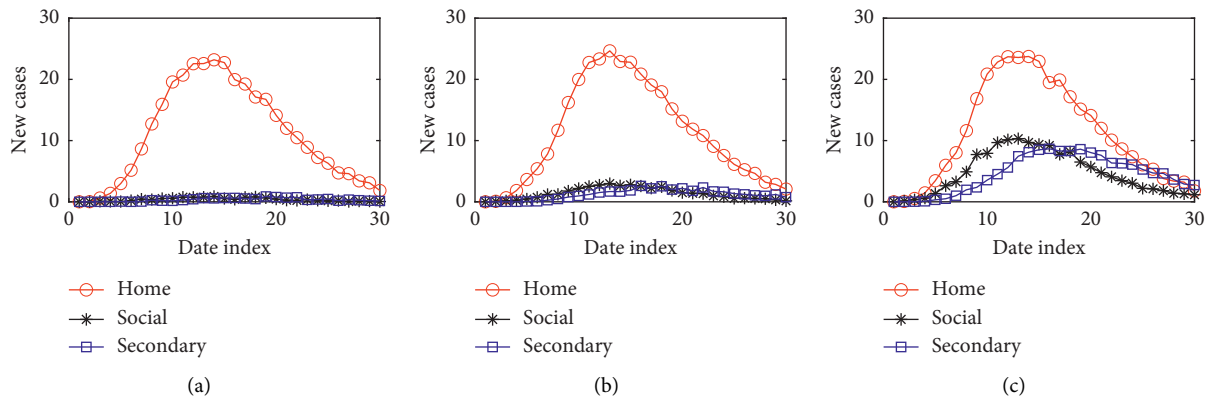


FIGURE 7: Continued.

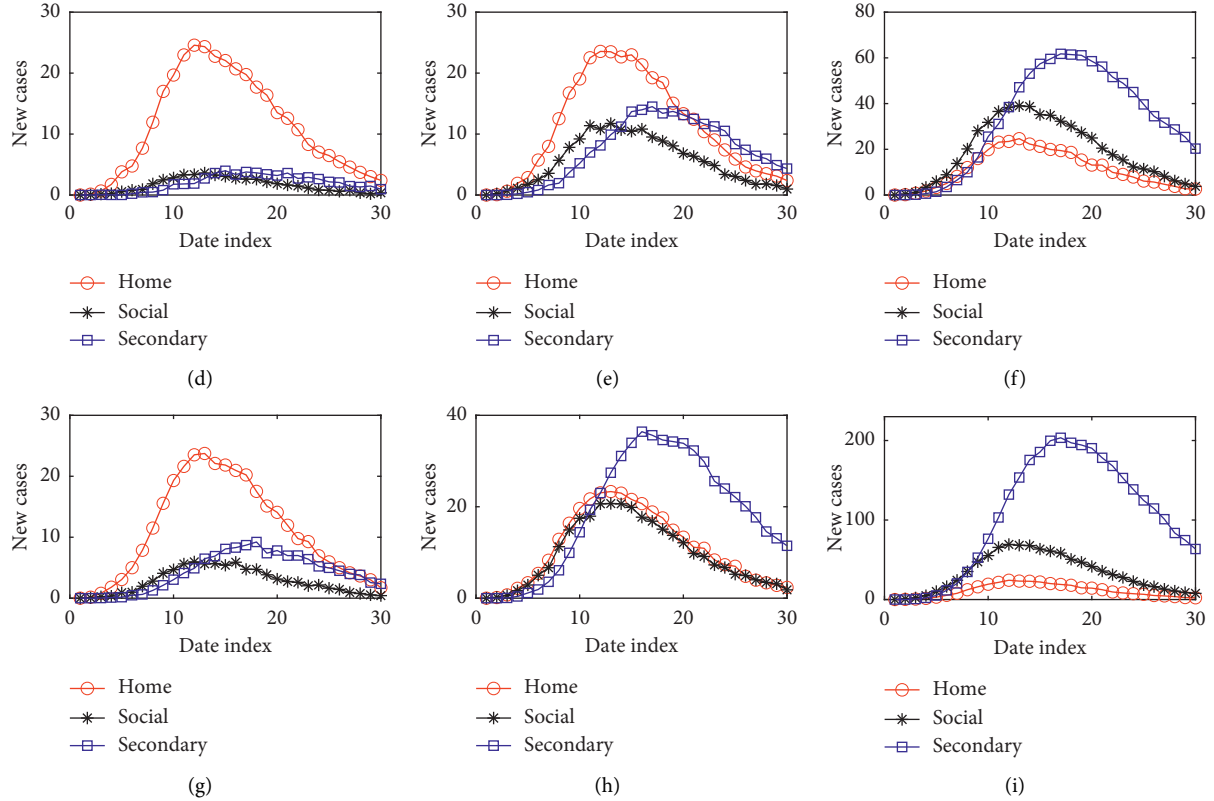


FIGURE 7: The simulation results of the age-dependent branching process with immigration. The red, blue, and black curves are for the local infectees divided as Home, Social, and Secondary. Home and Social refer to the infected cases of the imported cases taking place at home and out of home, respectively. Secondary refers to the infectees caused by Home and Social. The nine subfigures are drawn with different combinations of the values for the testing parameters as $\alpha = 0.1, 0.4, \text{ and } 0.7$ and $\lambda = 0.4, 1.4, \text{ and } 4.8$. (a) $\alpha = 0.1$ and $\lambda = 0.4$. (b) $\alpha = 0.1$ and $\lambda = 1.4$. (c) $\alpha = 0.1$ and $\lambda = 4.8$. (d) $\alpha = 0.4$ and $\lambda = 0.4$. (e) $\alpha = 0.4$ and $\lambda = 1.4$. (f) $\alpha = 0.4$ and $\lambda = 4.8$. (g) $\alpha = 0.7$ and $\lambda = 0.4$. (h) $\alpha = 0.7$ and $\lambda = 1.4$. (i) $\alpha = 0.7$ and $\lambda = 4.8$.

probability of going out is 70%; then, the local infectees grow very fast as λ increases from 0.4 and 1.4 to 4.8. In Figures 7(g)–7(i), the Home infectees keep stable, while the Social and Secondary increase obviously as λ increases. Notably, our assumption on contact tracing leads to a complete cutoff of the third and further generations. However, with $\alpha = 0.7$ and $\lambda = 4.8$, the number of infectious individuals is so large that it is quite difficult to isolate the infected individuals, let alone the trace back and isolation of the close contact individuals, due to the lack of medical resources. Therefore, an outbreak would take place in local communities with high possibility.

To sum up, when faced with the pandemic of COVID-19, the most costless and effective measure is staying at home. It is not the effort of someone but the effort of everyone. More importantly, it should be carried out simultaneously. However, considering the trade-off between the prevention of COVID-19 and economic affairs, keeping a proper social distance is more important.

5. Conclusion

Based on the confirmed cases reported outside the epic center in China, temporal and structural patterns are extracted from the actual data. Moreover, an age-dependent branching process with immigration is built to mimic the mechanism of the

transmission of COVID-19 in particular local communities. Our model matches the actual data quite well, showing the validation of our branching model. The efficiencies of isolation and social distance are also tested by the branching model. We reveal that the spreading chain can be cut efficiently under strict isolation, which might be the main reason for the success of COVID-19 prevention in China. However, due to the trade-off between economic consideration and prevention of the pandemic, keeping a proper social distance is more important when leaving home for social activities. Our findings reveal the effectiveness of isolation in China outside Hubei Province and may shed light on preventing the global pandemic spreading of COVID-19.

The branching structure is proper for modeling the spreading of COVID-19, as shown in Figure 4. Although the situations we considered are the well-prevented local communities, the basic features, such as the serial interval, the composition of infectees, and the immigration structures, can be applied to more general situations for investigating other containment measures.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was jointly supported by the National Natural Science Foundation of China (Grant nos. 11775034, 11971074, 11905042, and 61976025) and Fundamental Research Funds for the Central Universities (no. 2019XD-A11).

References

- [1] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020.
- [2] A. Lee, "Wuhan novel coronavirus (COVID-19): why global control is challenging?" *Public Health*, vol. 179, pp. A1–A2, 2020.
- [3] J. F.-W. Chan, S. Yuan, K.-H. Kok et al., "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *The Lancet*, vol. 395, no. 10223, pp. 514–523, 2020.
- [4] T. Zhou, Q. Liu, Z. Yang et al., "Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV," *Journal of Evidence-Based Medicine*, vol. 13, no. 1, pp. 3–7, 2020.
- [5] S. Zhao, Q. Lin, J. Ran et al., "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak," *International Journal of Infectious Diseases*, vol. 92, pp. 214–217, 2020.
- [6] D. Chen and T. Zhou, "Control efficacy on COVID-19," 2020, <http://arXiv.org/abs/2003.00305>.
- [7] Epidemiology Working Group for NCIP Epidemic Response Chinese Center for Disease Control and Prevention, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China," *Zhonghua Liu Xing Bing Xue Za Zhi*, vol. 41, pp. 145–151, 2020.
- [8] W. Guan, Z. Ni, Y. Hu et al., "Clinical characteristics of coronavirus disease 2019 in China," *NEJM*, vol. 382, pp. 1708–1720, 2020.
- [9] Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, and L. A. Meyers, "Serial interval of COVID-19 among publicly reported confirmed cases," *Emerging Infectious Diseases*, vol. 26, 2020.
- [10] Z. Yu, G. Zhang, Q. Liu, and Z. Lv, "The outbreak assessment and prediction of 2019-nCoV based on time-varying SIR model," *University of Electronic Science and Technology of China*, vol. 49, no. 3, pp. 357–361, 2020.
- [11] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China," *Science*, vol. 368, no. 6492, pp. 742–746, 2020.
- [12] L. Hien, K. Veria, P. Kocbach et al., "The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China," *Journal of Travel Medicine*, vol. 27, no. 3, Article ID taaa037, 2020.
- [13] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study," *The Lancet*, vol. 395, no. 10225, pp. 689–697, 2020.
- [14] G. Chowell, C. Viboud, L. Simonsen, and S. M. Moghadas, "Characterizing the reproduction number of epidemics with early subexponential growth dynamics," *Journal of The Royal Society Interface*, vol. 13, no. 123, Article ID 20160659, 2016.
- [15] J. Hellewell, S. Abbott, A. Gimma et al., "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts," *Lancet Glob Health*, vol. 8, no. 4, pp. 488–496, 2020.
- [16] L. Zhang, "Fitness of the generalized growth to the 2019 novel coronavirus data," *University of Electronic Science and Technology of China*, vol. 49, no. 3, pp. 345–348, 2020.
- [17] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez et al., "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, vol. 368, no. 6490, pp. 493–497, 2020.
- [18] Q. Li, W. Feng, and Y.-H. Quan, "Trend and forecasting of the COVID-19 outbreak in China," *Journal of Infection*, vol. 80, no. 4, pp. 469–496, 2020.
- [19] F. Ball and P. Donnelly, "Strong approximations for epidemic models," *Stochastic Processes and Their Applications*, vol. 55, no. 1, pp. 1–21, 1995.
- [20] J.-F. Wang, T.-L. Zhang, and B.-J. Fu, "A measure of spatial stratified heterogeneity," *Ecological Indicators*, vol. 67, pp. 250–256, 2016.

Research Article

Understanding the Anticontagion Process and Reopening of China during COVID-19 via Coevolution Network of Epidemic and Awareness

Lingbo Li , Ying Fan , An Zeng , and Zengru Di

School of Systems Science, Beijing Normal University, Beijing 100875, China

Correspondence should be addressed to Ying Fan; yfan@bnu.edu.cn and An Zeng; anzeng@bnu.edu.cn

Received 14 October 2020; Revised 29 December 2020; Accepted 16 April 2021; Published 11 May 2021

Academic Editor: Ning Cai

Copyright © 2021 Lingbo Li 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 novel coronavirus (COVID-19) pandemic is intensifying all over the world, but some countries, including China, have developed extensive and successful experience in controlling this pandemic. In this context, some questions arise naturally: What can countries caught up in the epidemic learn from China's experience? In regions where the outbreak is under control, what would lead to a resurgence of the epidemic? To address these issues, we investigate China's experience in anticontagion interventions and reopening process, focusing on the coevolution of epidemic and awareness during COVID-19 outbreak. Through an empirical analysis based on large-scale data and simulation based on a metapopulation and multilayer network model, we ascertain the impact of human movements and awareness diffusion on the epidemic, elucidate the inherent patterns and effective interventions of different epidemic prevention methods, and highlight the crunch time of each measure. The results are also employed to analyze COVID-19 evolution in other countries so as to find unified rules in complex situations around the world and provide advice on anticontagion and reopening policies. Our findings explain some key mechanisms of epidemic prevention and may help the epidemic analysis and decision-making in various countries suffering from COVID-19.

1. Introduction

Catching the world by surprise, the novel coronavirus (COVID-19) pandemic has hit over 210 countries and regions, affected more than seven billion people on the planet, and claimed over 480,000 precious lives [1]. Countries around the globe are experiencing different stages of the epidemic; e.g., the ongoing pandemic in Italy is almost under control, while Brazil is going through an outbreak [2, 3]. As one of the earliest countries challenged by COVID-19, China effectively mitigated the spread of this novel disease, shortened the duration of new cases increasing precipitously, and took appropriate action to promote economic recovery while ensuring the epidemic curve leveling off [4, 5]. In the battle against COVID-19, China has built much experience, from lockdown restrictions to awareness reminders, from anticontagion policies to reopening measures [6–8]. However, three

months after the epidemic was stably controlled in China, new cases in Beijing grew significantly again in mid-June. In this context, some questions arise naturally: Why could China be successful in stopping the pandemic? What led to the recent resurgence of the epidemic in Beijing?

A team organized by the World Health Organization (WHO) gives an unequivocal report of the pandemic in China, suggesting that China's bold approach has changed the course of a rapidly escalating and deadly epidemic and hopes China's successes will encourage other countries to act [9,10]. To discuss the experience of China and the relevance between policies in various countries, we investigate the impact of some epidemic factors. We highlight the key regions and crunch time, discuss the patterns and effective measures of the outbreak, and provide advice based on China's experience, which may benefit some parts of the world suffering from COVID-19.

To analyze COVID-19, some research builds epidemic models to simulate its spread between individuals [11–13] or utilizes statistical and econometric methods to empirically evaluate the impact of policies on the growth rate of infections [14, 15]. However, previous works neglected the coupling effect between multiple spreading processes in the epidemic. Actually, in complex social systems, diffusion usually does not occur independently but is coupled with each other and evolves together [16–18]. In the case of COVID-19, as the virus spreads among the crowd, information about precautionary measures also circulates simultaneously. These spreading processes affect each other and will both be promoted when the state goes back to work from anti-epidemic status. Therefore, it is necessary to understand the interaction of multiple dynamic processes and provide comprehensive solutions for both epidemic prevention and economic recovery.

This work focuses on the coevolution of epidemic spreading and awareness diffusion, taking into account the heterogeneous population mobility, imported cases, and other practical factors. It makes empirical analysis and develops a metapopulation and multilayer network model that leverages lots of real-time mobility and case data. We find evidence of the critical time point for each anti-contagion method, e.g., movement restriction and awareness reminder. We also summarize effective interventions that have been proved to achieve large and beneficial health outcomes in China. These findings may benefit the epidemic analysis and decision-making in the rest part of the world. The rest of this paper is organized as follows. In Section 2, we introduce the data resources and interpret our network model and its operating mechanism. In Section 3, we elucidate the results based on simulation and empirical analysis. More discussion can be found in Section 4.

2. Data and Model

2.1. Data Description. We carry out the research with large amounts of fine-grained real data, including (i) the registered population and resident population of each prefecture-level administrative area, extracted from the 2019 statistical yearbook of each province in China (<http://www.stats.gov.cn/tjsj/>); (ii) the passenger volume carried by modes of transportation, i.e., waterway, highway, railway, and civil aviation, obtained from the statistics released by the Ministry of Transport (<http://www.mot.gov.cn/>); (iii) the population migration index, migration destination, and local travel intensity of each prefecture-level administrative area, coming from the Baidu Migration (<https://qianxi.baidu.com/2020/>); (iv) the cumulative number of confirmed cases and the number of imported cases, derived from the Health Commissions (http://www.nhc.gov.cn/xcs/xxgzbd/gzbd_index.shtml); (v) measures and timing of epidemic prevention and reopening policy on a country-by-country basis. These data are used in simulation and verification, as well as empirical analysis.

2.2. Model. Epidemic models aim to depict the spread of diseases among people, which can be divided into SIS, SIR, SEIR, and other models according to the characteristics of the epidemic [19–23]. The outbreak of COVID-19 in China has the following properties:

- (i) Virus carriers have an incubation period of up to two weeks
- (ii) At the early stage of the epidemic, the population movement within and between regions amplified localized outbreaks of disease into widespread epidemics, whereas in the latter stage, the new cases increasing mainly results from the imported cases
- (iii) In the process of epidemic prevention, there is a continuous and high-intensity reminder of self-protection awareness and knowledge

The above factors also play an important role in the COVID-19 outbreak in various countries.

Based on these facts, we develop an aggregate model to reproduce the COVID-19 outbreak, which combines a metapopulation network and two-layer coupling networks (Figure 1). This model includes multiple practical factors, i.e., heterogeneous human movements, disease spreading, awareness diffusion, and case importation from abroad. We first establish a metapopulation network to describe the human migration between regions [24–26]. The network includes 345 nodes, i.e., 345 subpopulations, and each node represents a prefecture-level administrative unit in China (Figure 1(a)). The edges depict the individual diffusion between administrative units, and the weights of edges are described by the volume of real population movements between regions at high resolution. Hence, this network draws macroscenarios of the spatiotemporal distribution of COVID-19 in China.

Secondly, a two-layer coupling network is built for each subpopulation to portray the coevolution of epidemic and awareness within each city (Figure 1(b)), where nodes represent individuals [17, 27, 28]. The virtual contact layer describes the awareness diffusion based on the UAU model, and each node in this layer has two possible states: Unaware (U), representing that the individual has no self-protection awareness due to objective or subjective reasons, and Aware (A), referring to individuals with awareness of protecting themselves from this disease [29]. By contrast, the real contact layer supports the epidemic spreading, where each node may be one of the following four states: Susceptible (S), Exposed (E), Infected (I), or Removed (R) according to the classic SEIR model [30]. There is a one-to-one correspondence of nodes between the two layers. The real state of each node is the combined result, so the coupling network consists of up to seven kinds of nodes $Z_i = [US_i, UE_i, UR_i, AS_i, AE_i, AI_i, AR_i]$, e.g., aware infected node (AI) and unaware susceptible node (US). Here, we consider that once an individual becomes infected, he or she will not ignore this disease anymore, which indicates the absence of UI state [31]. Edges in the virtual contact layer represent conscious interactions caused by online or offline communication between individuals, while edges in the real contact layer represent infections caused by actual contact

between individuals. The coupling effect is reflected by the awareness-dependent exposure probability, which means that aware people (β^A) are less likely than unaware people (β^U) to be exposed to dangerous infectious environments, i.e., $\beta^A = \gamma * \beta^U$ ($\gamma < 1$).

Based on the model described above, the simulation experiments run as follows [25], with each round of evolution including two steps:

- (1) In the first stage, we count the individual diffusion between subpopulations and the overseas imports. Then, we update the total population and proportion of individuals in each state for every subpopulation; e.g., the total number of residents in the i^{th} subpopulation at $(t + 1)^{th}$ time step is $Z_i(t + 1) = Z_i(t) + \Delta Z_i(t)$.
- (2) In the second stage, epidemic and awareness evolve together within subpopulations. We first calculate the effective exposure rate as follows:

$$\begin{aligned}
 P_i^{US}(t + 1) &= [P_i^{US}(t) * (1 - \lambda) + P_i^{AS}(t) * \delta] * (1 - \beta^U) + [P_i^{UR}(t) * (1 - \lambda) + P_i^{AR} * \delta] * \eta, \\
 P_i^{UE}(t + 1) &= [P_i^{US}(t) * (1 - \lambda) + P_i^{AS}(t) * \delta] * \beta^U + [P_i^{UE}(t) * (1 - \lambda) + P_i^{AE} * \delta] * (1 - \alpha), \\
 P_i^{UR}(t + 1) &= [P_i^{UR}(t) * (1 - \lambda) + P_i^{AR}(t) * \delta] * (1 - \eta) + P_i^{AI}(t) * \mu * \delta, \\
 P_i^{AS}(t + 1) &= [P_i^{US}(t) * \lambda + P_i^{AS}(t) * (1 - \delta)] * (1 - \beta^A) + [P_i^{UR}(t) * \lambda + P_i^{AR} * (1 - \delta)] * \eta, \\
 P_i^{AE}(t + 1) &= [P_i^{US}(t) * \lambda + P_i^{AS}(t) * (1 - \delta)] * \beta^A + [P_i^{UR}(t) * \lambda + P_i^{AR} * (1 - \delta)] * (1 - \alpha), \\
 P_i^{AI}(t + 1) &= (P_i^{UE}(t) + P_i^{AE}(t)) * \alpha + P_i^{AI}(t) * (1 - \mu), \\
 P_i^{AR}(t + 1) &= [P_i^{UR}(t) * \lambda + P_i^{AR}(t) * (1 - \delta)] * (1 - \eta) + P_i^{AI}(t) * (1 - \delta) * \mu,
 \end{aligned} \tag{2}$$

where λ is the awareness diffusion rate, δ is the awareness forgetting rate, α is the infection rate, μ is the removal rate, and η is the reinfection rate. These probabilities can be utilized to calculate the population of each state in the i^{th} subpopulation at the $(t + 1)^{th}$ time step.

3. Results

3.1. Epidemic Simulation. By employing our model to simulate the outbreak, we reproduced the spread of COVID-19 in China from January 24th (i.e., the Chinese New Year's Eve) to mid-April. The parameter set, which makes the simulation result closest to the real final cumulative number of cases, is selected, such that the epidemic is therefore explained well by the simulation. We extract the difference between the simulation results and real data of each province (except Hong Kong, Macao, and Taiwan) for further analysis (Figure 2(a)).

These differences suggest a segregation effect since the results in eastern and southwestern China differ significantly. Hence, we map the data to the real geographic location (Figure 2(b)). According to the map, we take the antiepidemic achievements of some municipalities as the benchmark, i.e., Beijing, Tianjin, and Shanghai. It is found

$$\beta_i^{eff}(t) = \text{inRate} * \beta * \frac{N_i^{AI}(t)}{Z_i(t)}, \tag{1}$$

where inRate is the local travel intensity within regions, β is the basic exposure rate, and $N_i^{AI}(t)$ and $Z_i(t)$ are the number of infected (AI) and total population in the i^{th} subpopulation at t^{th} time step, respectively. Therefore, $\text{inRate} * (N_i^{AI}(t)/Z_i(t))$ indicates an individual's possible exposure to infected individuals. Given that β is the exposed probability of an individual in a single contact with the infected population, $\beta_i^{eff}(t)$ depicts the actual exposed probability of an individual in possible several contacts with the infected population of the i^{th} subpopulation at t^{th} time step. Next, the probability of an individual being in a certain state at $(t + 1)^{th}$ time step is updated as follows:

that the severity of the epidemic in the central region and east coast, as well as the Heilongjiang province, is relatively high (the red region in Figure 2(b), representing the places where the simulation results are lower than the real data). Nevertheless, the severity of the epidemic in the southwest region is relatively low (the blue region in Figure 2(b), representing the places where the simulation results are higher than the real data).

To some extent, this phenomenon indicates that the spread of COVID-19 is lower than expected in southwest China while higher in central and eastern China. This might be due to multiple reasons. On the one hand, given the low population density in the western and southern regions, the epidemic spreading is relatively difficult; on the other hand, regions with frequent overseas trade and population transfer, e.g., the east coast and Heilongjiang, are particularly affected by the international outbreak.

3.2. Key Factors Analysis. In the model setup, we consider the impact of human movement (including population mobility within cities and the population migration between cities) and awareness diffusion on the epidemic. The Baidu Migration website provides population mobility indexes to show the volume of population movement, and the

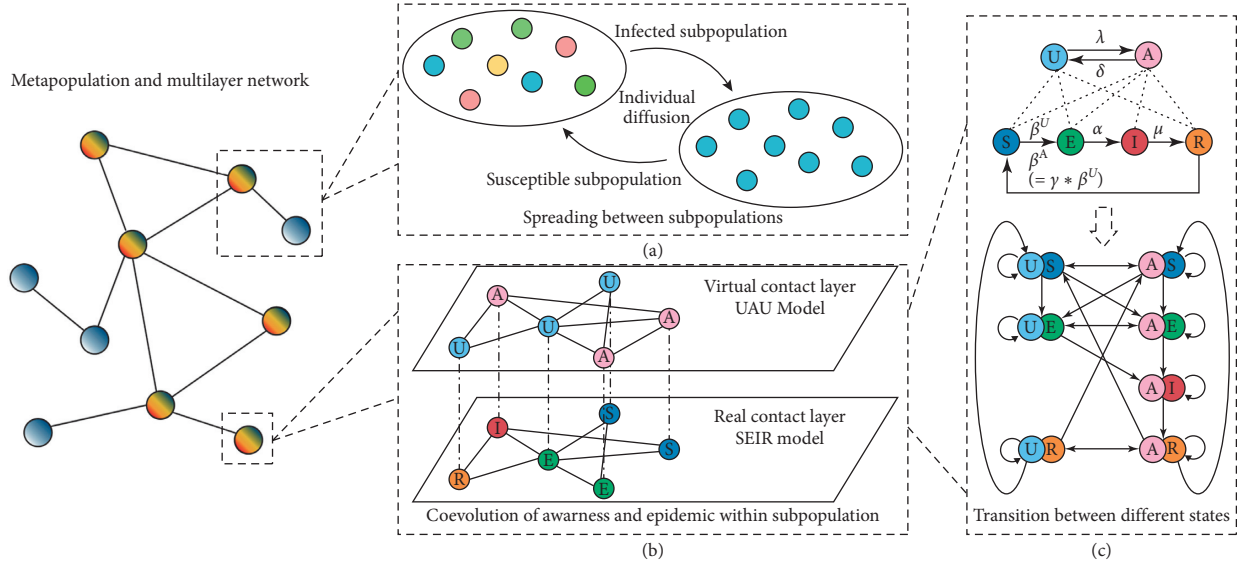


FIGURE 1: A simple illustration of our model framework. The left part displays the metapopulation network, where the blue and rainbow-colored nodes represent the susceptible and infected subpopulations, respectively. Subfigure (a) represents the individuals of each subpopulation moving between different subpopulations along with the epidemic and awareness. Subfigure (b) demonstrates the awareness diffusion (UAU model) and epidemic spreading (SEIR model) interaction based on virtual or real individual contacts within each subpopulation, modeled by a multilayer coevolution network. (c) indicates the mechanism of transition between different states based on several parameters. (a) Spreading between subpopulation. (b) Coevolution of awareness and epidemic within a subpopulation. (c) The transition between different states.

comparison of these indexes indicates the change of population flow in different regions or periods. We adjust these key factors and try to understand how these factors affect epidemic prevention. To be specific, we make the following adjustments to the three factors: (1) global adjustment for all regions and periods; (2) local adjustment, especially adjustments in Hubei province; (3) temporal adjustment in the early, middle, and later periods.

3.2.1. Data Analysis and Impact of Factors

(1) *Factor One: Population Mobility within Cities.* First, comparing real-time mobility data on a year-by-year basis can illustrate how the population mobility within each city is controlled. The data suggest that the average population mobility within cities was limited to 81.3% of the same period last year. There is a power-law exponent of around 0.96 between the 2020 and 2019 population mobility indexes. Regions with significantly strong restrictions include Hubei and Xinjiang provinces, which were controlled at around 58.8% (Figure 3(a)). Next, we adjust the actual population mobility data in China from the above three aspects: global control, local control in Hubei province, and temporal control (Figure 3(b)). Results indicate that global control causes the most dramatic changes in the epidemic, followed by the control in Hubei. Both of these controls lead to an exponential change in the cumulative cases. It can be found that the population movement control within Hubei, the worst-hit area in China, can produce an important effect on the evolution of COVID-19.

(2) *Factor Two: Population Migration between Cities.* Similarly, we identify that the average population migration

between cities was limited to 60.6% of the same period last year. The power-law exponent between the 2020 and 2019 population migration indexes is about 0.91. Beijing, Hong Kong, and Hubei provinces imposed strict restrictions, limiting the migration to between 22.2% and 45.4% (Figure 3(c)). This evidence elucidates that regional governments have taken much tougher measures to control the migration between cities than mobility within cities. Results also show that the control at the early stage has a comparable influence on the epidemic with the global control (Figure 3(d)). This phenomenon underlines the necessity and effectiveness of early intervention in the management of population migration between cities.

(3) *Factor Three: Awareness Diffusion.* We study three model parameters related to self-protection awareness, i.e., the rate of awareness diffusion, the rate of awareness forgetting, and the proportion of awareness in the initial configuration. The experimental results suggest that adjusting the awareness diffusion rate makes the greatest impact on the epidemic, compared with the other two (Figure 3(e)). It can be judged by the parameter value that self-protection awareness has been widely spread during the COVID-19 outbreak in China. The awareness diffusion rate is selected as the main variable for controlling the awareness spreading. We find that the earlier we take action, the more remarkable the change in the epidemic will be (Figure 3(f)). This phenomenon also applies to the previous two factors. The result of awareness ascertains the important role played by self-protection awareness, which is created by publicity and information dissemination, in the war against the pandemic.

(4) *Comparison of Factors.* We further make a comparative analysis of the above three factors. On the basis of China's

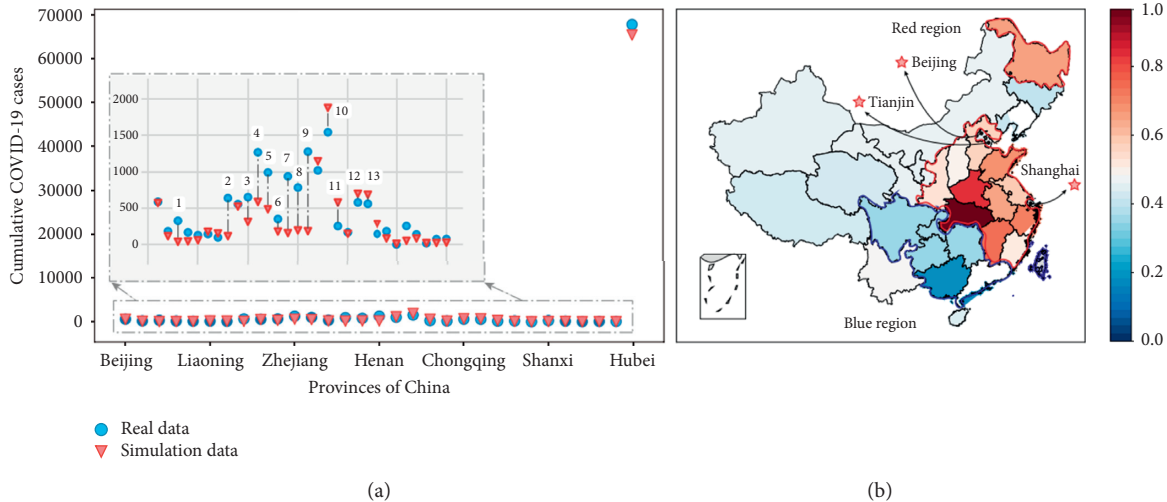


FIGURE 2: Comparison of simulation results and real epidemic data in China. In subfigure (a), the blue circles and red triangles represent the real and simulated cumulative numbers of confirmed cases in Chinese provinces on April 12th, respectively, with 31 provinces in total ranked by administrative code (except Hong Kong, Macao, and Taiwan). A partial enlarged view of the dotted rectangle highlights the data from 30 provinces except Hubei. In subfigure (b), the differences between real and simulated cumulative confirmed cases were converted into a heat map to observe the geographical distribution of the epidemic. The red region indicates that the real data is higher than the simulation result, while the blue region indicates the opposite. (a) Simulation results vs. real data. (b) Distribution in India.

practical measures, it brings about the most dramatic change in the epidemic to adjust the population mobility with cities, followed by awareness spreading and population migration between cities (Figures 3(b)–3(f)).

On the one hand, the results confirm the lockdown restrictions between regions in China. Actually, China attached great importance to avoiding the hidden risk of the epidemic transferring among cities and strictly confined the human movement between cities. Therefore, the reopening policy began within cities, and college students were prohibited from returning to school to avoid mass migration across regions. On the other hand, the results emphasize the importance of controlling population mobility within cities and anticontagion awareness. The human activities within each region in China were limited by mandatory stay-at-home order, closure of physical business premises, and cancellation of nonessential gatherings. China also conducted anticontagion publicity through various channels, e.g., mainstream media, We media, and slogans, to improve the awareness of self-protection and knowledge of residents.

We also study the effect of factor adjustment in key areas, i.e., Hubei province in this case, which is the worst-hit area in China. Results show that the population mobility within Hubei leads to greater changes in the epidemic than population migration from this province (Figure 4(a)). This indicates that there was room for improving the human movement control within Hubei, whereas the population outflow from Hubei has been fully restricted, which is confirmed by actual data (Figure 4(b)). Therefore, it is reasonable to believe that if the population movement in Hubei is more strictly confined during the outbreak, the epidemic in China will be more effectively mitigated.

On the whole, we adjust the three key factors by global, local, and temporal adjustments. Results show the crucial

influence of these factors on the epidemic. Although many people emphasize the travel restriction between regions, equal attention should be given to population mobility control within each region and awareness reminders for the residents. We highlight the importance of measures in Hubei, especially controlling the population mobility within this region, and evaluate the epidemic prevention in Hubei through empirical analysis. We also find that early interventions always do more with less for each factor. Hence, we next conduct more analysis to discover how measures in different stages affect the epidemic and what effective measures and crunch time exist.

3.2.2. Anticontagion and Reopening Policies. In the light of the above findings, we further examine the three factors in detail to discuss specific crunch time and effective measures. We focus on the impacts of them on the epidemic at different periods and for different goals, i.e., epidemic prevention and economic recovery. Since we regard each week as a period, the simulation duration from late January to mid-April is divided into 11 periods. The results reveal that the control of each factor has a critical time point. If we do not take the crunch time seriously, it is highly likely that a large-scale outbreak of pandemic will occur.

(1) Factor One: Population Mobility within Cities. We first simulate the reopening timetable by implementing the resumption policy from a certain period in turn. It is shown that if work resumes within cities earlier than the fifth week when the new cases almost decay to zero, the previous anticontagion achievements will be offset, leading to an outbreak (Figure 5(a)). Actually, China went back to work from around February 23th, which is just in the fifth week. This timing was exactly consistent with the simulation result.

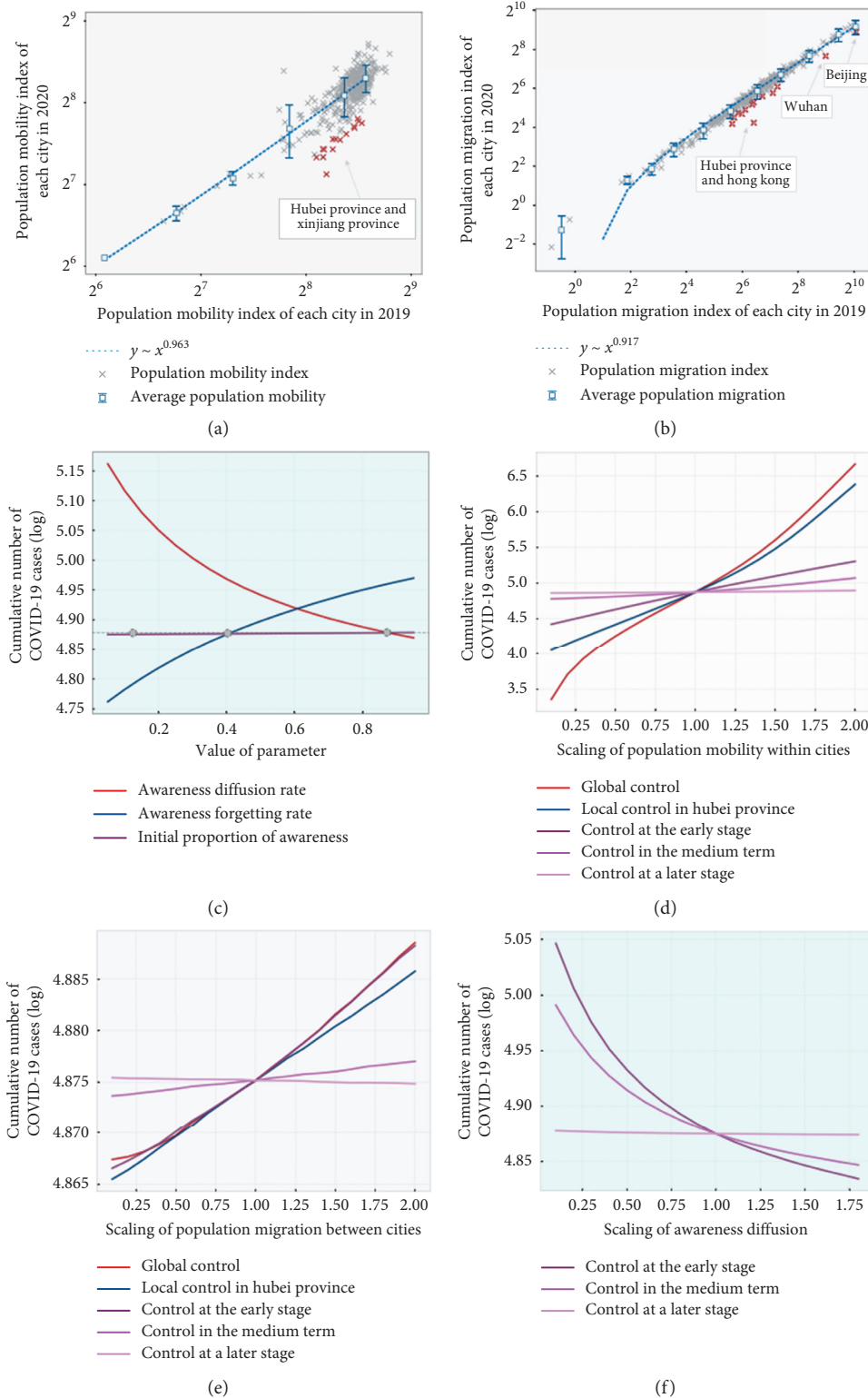


FIGURE 3: Real data analysis and the impact of three key epidemic factors, i.e., population mobility within cities, population migration between cities, and awareness diffusion. Subfigure (a) demonstrates the comparison between the 2019 and 2020 population mobility index of each city with a fitted parameter of 0.96. Subfigure (b) displays that the simulated cumulative confirmed cases change along with the control measures of population mobility within cities. The abscissa indicates the scaling factors. That is, the actual volume is used as the standard to zoom in or out. We perform the global, local, and temporal controls on an actual basis, with scaling ratios ranging from 0.1 to 2.0. Subfigures (c) and (d) show the results of controlling the population migration between cities in the same way. Subfigure (e) reveals the impact of three model parameters affecting awareness diffusion. The grey dots record the value of each parameter used in the model. Subfigure (f) demonstrates the outcomes of applying temporal control to awareness diffusion. (a) Data of population mobility within cities. (b) Data of population migration between cities. (c) Parameter of awareness diffusion. (d) Impact of population mobility within cities. (e) Impact of population migration between cities. (f) Impact of awareness diffusion.

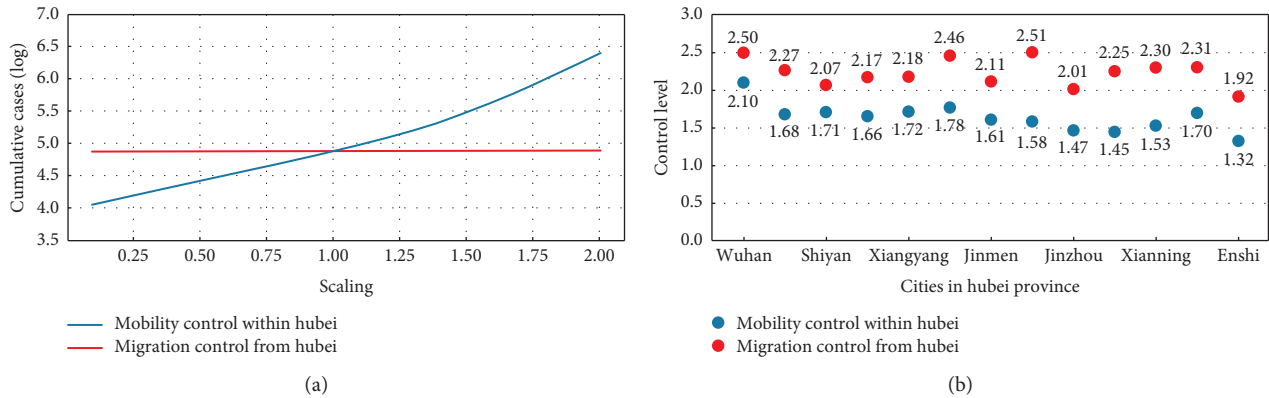


FIGURE 4: Comparison between the impacts of different controls in Hubei province. Subfigure (a) provides the simulation results of controlling the population mobility within Hubei (blue line) and migration from Hubei (red line) on a practical basis, with scaling ratios ranging from 0.1 to 2.0. In subfigure (b), the blue and red dots represent the real control level of population mobility within Hubei and migration from Hubei, respectively, which is calculated as dividing the 2019 average index by the 2020 average index of the same period. The real data identifies the much more strict restriction on population migration from Hubei than mobility within Hubei, which is consistent with the simulation results. (a) The simulation results of Hubei province. (b) Real data of Hubei province.

Therefore, the reopening within cities of China is reasonable and has effectively promoted economic recovery and epidemic prevention. Next, the epidemic prevention timetable shows the results of restricting human mobility within cities from a certain period in turn. We find the critical time appears in the fourth week, and if the population mobility restriction misses this time point, the intervention will have little effect (Figure 5(a)).

In addition, we investigate the impact of stricter prevention, which means taking tougher control measures on human movement within cities than the actual policy from a certain period in turn. As a result, only the strict restriction started within the first three weeks is efficient at slowing the growth substantially (Figure 5(a)). If we aim to reduce the final cumulative confirmed cases in China to less than half of the real data (i.e., less than 40,000) by controlling the population mobility within cities, the action should be taken in the first two weeks, i.e., limiting it to 76% of the current control level from the first week, or 64% from the second week (Figure 5(b)).

(2) *Factor Two: Population Migration between Cities.* The analysis of migration between cities is conducted in the same way as above. We find that the critical time points of deploying and lifting the migration restrictions between cities are the third and second week, respectively (Figure 5(c)). It means that we should take action to restrict the migration across regions at least before the explosive growth of new cases and restart it after the new cases drop to half the peak. Since these time points are earlier than the previous factor, it again supports the conclusion that we should pay more attention to the early control of migration between cities. Results also reveal that the stricter restriction on migration across regions should start from the first week to effectively mitigate the outbreak (Figure 5(c)). However, the unprecedented migration confined in China is almost sufficiently strict. Therefore, if only tightening it further, the final cumulative confirmed cases will not be cut by half (Figure 5(d)).

(3) *Factor Three: Awareness Diffusion.* Similarly, the experiments on awareness show that we should start an awareness reminder no later than the third week and wait until at least the fourth week before proceeding to relax the publicity of anticontagion awareness (Figure 5(e)). It reveals that we should always value the awareness reminder at least until the new cases decline to near zero. If we try to limit the epidemic more efficiently, more efforts need to be put into publicizing awareness for the first three weeks. By contrast, the decline of the epidemic caused by the awareness reminder is weaker than its increase, which means that the self-protection awareness of Chinese residents is relatively strong. On this basis, it is not achievable to bring the cumulative cases below 40,000 by only controlling awareness spreading (Figure 5(f)).

(4) *Summary of Comparison.* In general, population mobility within cities, migration between cities, and awareness diffusion are proven to be key epidemic factors. Taking appropriate measures from these aspects at the right time can effectively control the pandemic (Figures 5(a)–5(e)). The controlling of these factors changes the course of a rapidly escalating and deadly epidemic, and the crunch time for epidemic prevention is discussed above. We also obtain consistent evidence that countries should reopen at least after the new case drops to half the peak.

However, once a state reopens, these human activity and awareness factors may lead to a resurgence of the epidemic at any time. The sudden resurgence of COVID-19 in Beijing in mid-June has refocused attention on epidemic prevention and control. In the previous three months, the country gradually resumed work and classes, and large-scale gatherings and movements occurred again. The anticontagion awareness reminder to the public also slackened. As a result, these conditions provide an opportunity for the epidemic to spread. Fortunately, Beijing immediately stopped resuming work, restricted population movement, updated the public on the epidemic information, and conducted a large number

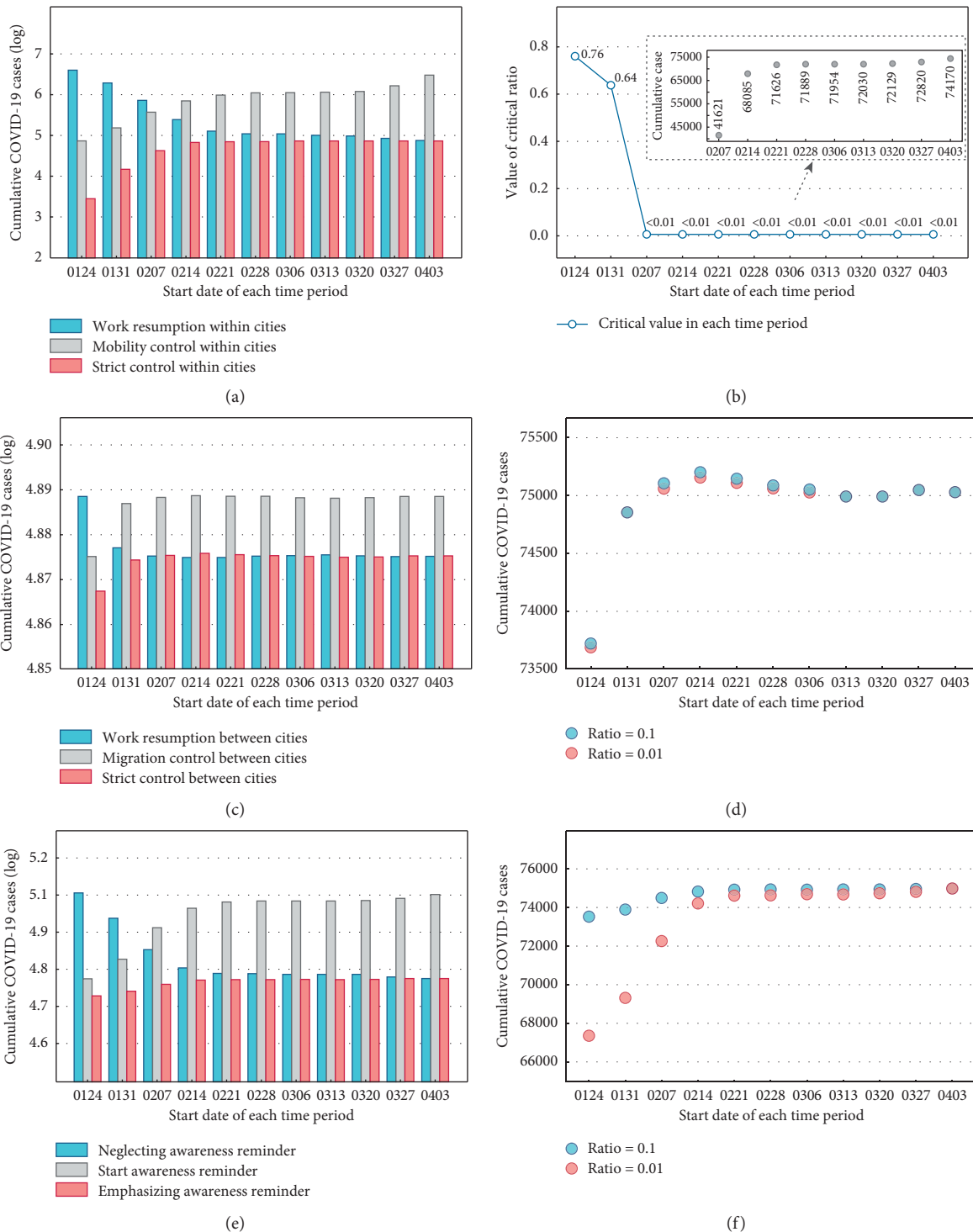


FIGURE 5: Timetable for the work resumption, epidemic prevention, and strict restrictions. (a) records the simulated cumulative confirmed cases, when the state goes back to work within each city (blue bar), restricts the population mobility within cities (grey bar), and takes tougher control measures on mobility within cities (red bar) since different periods. (c) shows the results of controlling the population migration between cities in the same way. The blue, grey, and red bars in (e) demonstrate neglecting, starting, and emphasizing the awareness reminder, respectively. To cut the final cumulative confirmed cases by half, (b) shows the critical limiting ratio of population mobility within cities. The partial enlarged view in (b) documents the cumulative cases when this factor is adjusted to 1% of the true level, which shows that the goal cannot be achieved in the third week or beyond. Since the goal to cut cumulative cases by half also cannot be achieved by only controlling the other two factors, i.e., the population migration between cities and awareness diffusion, (d) and (f) document the impact of limiting them to 1% and 10% of the true level. (a) Timeline for population mobility management. (b) Timeline for population migration management. (c) Timetable for awareness diffusion. (d) Critical value of population mobility management. (e) Critical value of population migration management. (f) Critical value of awareness diffusion.

of viral tests, which quickly curbs the growth of coronavirus cases.

3.3. Experience of China and Reality of Various Countries.

On the basis of the above analysis, we summarize the characteristics and experience of China in epidemic prevention and economic recovery:

- (1) The epidemic prevention started at the initial stage of the outbreak when new cases just began to increase exponentially. The key measures included restrictions on population mobility within cities, restrictions on population migration between cities, and publicity of anticontagion awareness.
- (2) The reopening policy was implemented when the pandemic almost was under control, i.e., new cases decreased almost to zero. The resumption of work started within cities and then extended to human movements across cities.
- (3) Large-scale new cases in China lasted about one month.

The war against COVID-19 of China has gained remarkable achievements. Based on China's experience in the epidemic prevention and reopening process, we analyze the epidemic policies in various countries, where we focus on the deploying and lifting of restrictions on population movement. The results are presented in Figure 6, and the main findings are listed as follows:

- (1) Some Asian countries, e.g., Japan and South Korea [32, 33], also started anticontagion measures at the beginning of the outbreak and resumed work when the epidemic is under stable control (Figure 6(a)). The outbreak was short-lived in these countries.
- (2) In some European countries, e.g., the United Kingdom, Italy, and Germany [34, 35], the epidemic prevention started relatively late, with a time lag of about two weeks from the initial outbreak, and the reopening measures took effect when the epidemic was initially under control (Figure 6(b)). For example, in the early stages of the pandemic, the UK adopted loose anticontagion measures, delayed lockdown restrictions, and the self-protection awareness of the residents was weak. These led to an insufficient implementation of the government's recommendations, such as keeping a social distance, which made the outbreak last longer.
- (3) In the American countries, e.g., the United States and Canada [36, 37], anticontagion policies were put forward in the early stage of the outbreak, but the practical measures were not strict enough. The countries got back to work when the outbreak was still in progress. Hence the epidemic lasted for a long time (Figure 6(c)). For instance, in the early days of

epidemic prevention in the United States, effective measures such as wearing masks were not emphasized enough by the government and many residents ignored personal protection, resulting in the explosive growth of coronavirus cases. In addition, many states began to lift social distancing measures and organize resumption of production even before the new cases showed a clear downward trend, and due to social causes, large-scale protest marches broke out in many parts of the United States. As a result, rather than being restricted, the population movement has become more active. These complex conditions make it difficult for the US epidemic to be effectively controlled.

Some other countries around the world now are still in epidemic growth, e.g., India, Russia, Mexico, and Brazil [38]. These countries began to prevent the epidemic early, but due to economic pressure and the world environment, they began to resume work during the explosive growth. Therefore, there is a risk of sustained large-scale outbreaks in these countries.

4. Discussion

This work investigates the coevolution of the epidemic spreading and awareness diffusion during the COVID-19 outbreak. It develops a network model with large-scale real data of heterogeneous human movements and imported cases to simulate the epidemic. We focus on three key factors of the epidemic, i.e., population mobility within regions, population migration between regions, and awareness diffusion, examining their effective measures and crunch time separately. Results elucidate the important role played by these factors in epidemic prevention and ascertain the impact of measures taken in Hubei on the national outbreak. We also highlight the critical time points for each anticontagion measure.

The experience of China is employed to analyze COVID-19 in other parts of the globe, to find unified rules in 275 complex situations around the world and give advice on epidemic prevention. The advice includes avoiding gathering and lack of self-protection awareness; strongly preventing the coronavirus from spreading within key regions or beyond; attaching importance to the crunch time, to eliminate risks at the source, and keeping the pandemic from recurring caused by reopening measures. Our findings explain some key mechanisms of epidemic prevention and contribute to the decision-making in various countries suffering from COVID-19.

We attempt to discuss the impact of awareness diffusion on the COVID-19 pandemic in this paper. Since awareness plays an important role in controlling epidemic spreading, many related issues are worth studying. We hope that this paper will inspire more work on the relationship between the epidemic and awareness and encourage more countries to

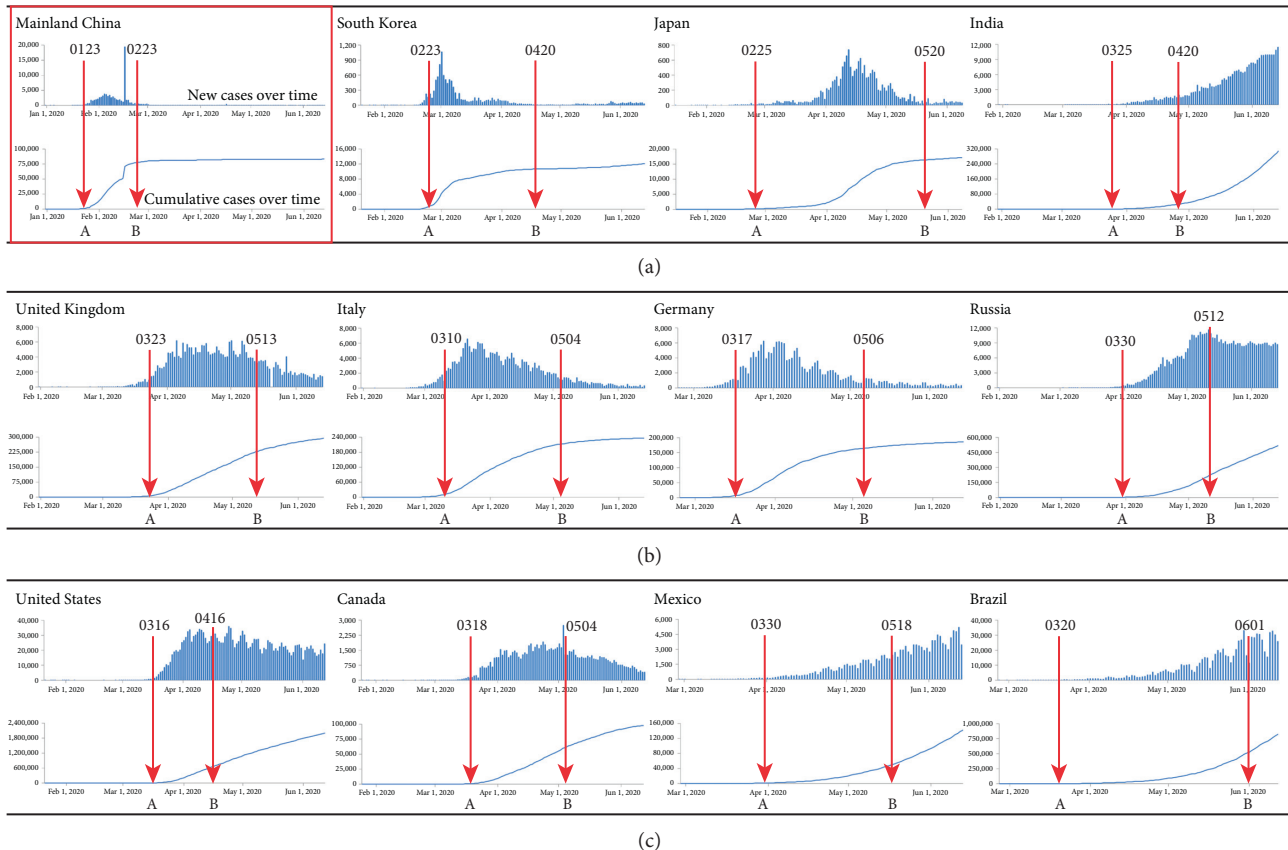


FIGURE 6: Anticontagion and reopening policies in various countries. We select 12 countries from Asia, Europe, and America as examples and mark the time points at which they took antiepidemic and reopening measures, corresponding to points A and B in the figure, respectively. We demonstrate the new cases and cumulative cases over time in the COVID-19 outbreak of each country. (a) Asia. (b) Europe. (c) America.

build residents' awareness of epidemic prevention in the battle against COVID-19.

Data Availability

The data source is described within the manuscript. The data supporting our results can be found in the official government websites of relevant countries and the following websites: <http://www.stats.gov.cn/tjsj/>, <http://www.mot.gov.cn/>, <https://qianxi.baidu.com/2020/>, and http://www.nhc.gov.cn/xcs/xgzbd/gzbd_index.shtml.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Conceptualization was performed by YF and AZ; formal analysis was carried out by LL, YF, and AZ; LL was responsible for providing software and visualization and original draft preparation. LL and AZ were responsible for editing and supervision.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 71731002).

References

- [1] I. Ali and M. L. Omar, "Covid-19: disease, management, treatment, and social impact," *Science of the Total Environment*, vol. 728, Article ID 138861, 2020.
- [2] M. Gatto, E. Bertuzzo, L. Mari et al., "Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures," in *Proceedings of the National Academy of Sciences*, vol. 117, no. 19, pp. 10484–10491, 2020.
- [3] W. Clare, J. Smith, and R. Morgan, "Covid-19: the gendered impacts of the outbreak," *The Lancet*, vol. 395, no. 10227, pp. 846–848, 2020.
- [4] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in recent confirmed covid-19 cases in China," *Science*, vol. 368, no. 6492, pp. 742–746, 2020.
- [5] Z. Wang, J. Wang, and J. He, "Active and effective measures for the care of patients with cancer during the covid-19 spread in China," *JAMA Oncology*, vol. 6, no. 5, pp. 631–632, 2020.
- [6] C. R. Wells, P. Sah, S. M. Moghadas et al., "Impact of international travel and border control measures on the global

- spread of the novel 2019 coronavirus outbreak,” in *Proceedings of the National Academy of Sciences*, vol. 117, no. 13, pp. 7504–7509, 2020.
- [7] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez et al., “The effect of human mobility and control measures on the covid-19 epidemic in China,” *Science*, vol. 368, no. 6490, pp. 493–497, 2020.
- [8] J. S. Jia, X. Lu, Y. Yuan, Ge Xu, J. Jia, and N. A. Christakis, “Population flow drives spatio-temporal distribution of covid-19 in China,” *Nature*, vol. 582, pp. 1–5, 2020.
- [9] K. Kupferschmidt and J. Cohen, “Can China’s covid-19 strategy work elsewhere?,” *Science*, vol. 367, no. 6482, pp. 1061–1062, 2020.
- [10] J. Gao, Y. Yin, B. F. Jones, and D. Wang, “Quantifying policy responses to a global emergency: insights from the covid-19 pandemic,” 2020, <https://arxiv.org/abs/2006.13853>.
- [11] M. Chinazzi, J. T. Davis, M. Ajelli et al., “The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak,” *Science*, vol. 368, no. 6489, pp. 395–400, 2020.
- [12] H. Tian, Y. Liu, Y. Li et al., “An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in China,” *Science*, vol. 368, no. 6491, pp. 638–642, 2020.
- [13] Y. Liu, T. W. Russell, J. K. Adam et al., “The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in Wuhan, China: a modelling study,” *The Lancet Public Health*, vol. 5, no. 5, pp. e261–e270, 2020.
- [14] H. Solomon, D. Allen, S. Annan-Phan et al., “The effect of large-scale anticontagion policies on the covid-19 pandemic,” *Nature*, vol. 1–9, 2020.
- [15] G. Bonaccorsi, F. Pierri, M. Cinelli et al., “Economic and social consequences of human mobility restrictions under covid-19,” in *Proceedings of the National Academy of Sciences*, vol. 117, no. 27, pp. 15530–15535, 2020.
- [16] W. Wang, Q.-H. Liu, J. Liang, Y. Hu, and T. Zhou, “Co-evolution spreading in complex networks,” *Physics Reports*, vol. 820, pp. 1–51, 2019.
- [17] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen, “The spread of awareness and its impact on epidemic outbreaks,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6872–6877, 2009.
- [18] J.-Q. Kan and H.-F. Zhang, “Effects of awareness diffusion and self-initiated awareness behavior on epidemic spreading - an approach based on multiplex networks,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 44, pp. 193–203, 2017.
- [19] H. Wang, Li Qian, G. D’Agostino, S. Havlin, H. E. Stanley, and P. Van Mieghem, “Effect of the interconnected network structure on the epidemic threshold,” *Physical Review E*, vol. 88, no. 2, Article ID 022801, 2013.
- [20] M. J. Keeling, T. House, A. J. Cooper, and L. Pellis, “Systematic approximations to susceptible-infectious-susceptible dynamics on networks,” *PLoS Computational Biology*, vol. 12, no. 12, Article ID e1005296, 2016.
- [21] N. Antulov-Fantulin, A. Lančić, H. Štefančić, and M. Šikić, “FastSIR algorithm: a fast algorithm for the simulation of the epidemic spread in large networks by using the susceptible-infected-recovered compartment model,” *Information Sciences*, vol. 239, pp. 226–240, 2013.
- [22] H.-F. Zhang and Z. Wang, “Suppressing epidemic spreading by imitating hub nodes’ strategy,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 10, pp. 1979–1983, 2019.
- [23] H.-F. Zhang, J.-R. Xie, M. Tang, and Y.-C. Lai, “Suppression of epidemic spreading in complex networks by local information based behavioral responses,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 24, no. 4, Article ID 043106, 2014.
- [24] E. A. Fronhofer and F. Altermatt, “Classical metapopulation dynamics and eco-evolutionary feedbacks in dendritic networks,” *Ecography*, vol. 40, no. 12, pp. 1455–1466, 2017.
- [25] S. Venkatramanan, J. Chen, A. Fadikar et al., “Optimizing spatial allocation of seasonal influenza vaccine under temporal constraints,” *PLoS Computational Biology*, vol. 15, no. 9, Article ID e1007111, 2019.
- [26] T. Jessica, P. Nicola, Q. Zhang, Y. Moreno, and A. Vespignani, “Phase transitions in information spreading on structured populations,” *Nature Physics*, vol. 16, no. 5, pp. 590–596, 2020.
- [27] H.-F. Zhang, J.-R. Xie, H.-S. Chen, C. Liu, and M. Small, “Impact of asymptomatic infection on coupled disease-behavior dynamics in complex networks,” *EPL (Europhysics Letters)*, vol. 114, no. 3, p. 38004, 2016.
- [28] J. Sun, L. Feng, J. Xie, Ma Xiao, D. Wang, and Y. Hu, “Revealing the predictability of intrinsic structure in complex networks,” *Nature Communications*, vol. 11, no. 1, pp. 1–10, 2020.
- [29] C. Xia, Z. Wang, C. Zheng et al., “A new coupled disease-awareness spreading model with mass media on multiplex networks,” *Information Sciences*, vol. 471, pp. 185–200, 2019.
- [30] 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.
- [31] C. Granell, S. Gomez, and A. Arenas, “Dynamical interplay between awareness and epidemic spreading in multiplex networks,” *Physical Review Letters*, vol. 111, no. 12, Article ID 128701, 2013.
- [32] E. Shim, A. Tariq, W. Choi, Y. Lee, and G. Chowell, “Transmission potential and severity of covid-19 in South Korea,” *International Journal of Infectious Diseases*, vol. 93, pp. 339–344, 2020.
- [33] K. Karako, P. Song, Y. Chen, and W. Tang, “Analysis of covid-19 infection spread in Japan based on stochastic transition model,” *Bioscience Trends*, vol. 14, no. 2, pp. 134–138, 2020.
- [34] L. Rosenbaum, “Facing covid-19 in Italy - ethics, logistics, and therapeutics on the epidemic’s front line,” *New England Journal of Medicine*, vol. 382, no. 20, pp. 1873–1875, 2020.
- [35] J. Yuan, M. Li, G. Lv, and Z. K. Lu, “Monitoring transmissibility and mortality of covid-19 in Europe,” *International Journal of Infectious Diseases*, vol. 95, pp. 311–315, 2020.
- [36] X. Marchand-Senécal, R. Kozak, S. Mubareka et al., “Diagnosis and management of first case of COVID-19 in Canada: lessons applied from SARS-CoV-1,” *Clinical Infectious Diseases*, vol. 71, no. 16, pp. 2207–2210, 2020.
- [37] M. Seyed, A. Shoukat, M. C. Fitzpatrick et al., “Projecting hospital utilization during the covid-19 outbreaks in the United States,” in *Proceedings of the National Academy of Sciences*, vol. 117, no. 16, pp. 9122–9126, 2020.
- [38] J. Rodriguez-Morales, V. Gallego, J. Escalera-Antezana et al., “Covid-19 in Latin America: the implications of the first confirmed case in Brazil,” *Travel Medicine and Infectious Disease*, vol. 35, Article ID 101613, 2020.

Review Article

Recent Progress about Flight Delay under Complex Network

Tang Zhixing,^{1,2} Huang Shan,³ and Han Songchen ²

¹College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan, China

²School of Aeronautics and Astronautics, Sichuan University, Chengdu, China

³Library, Civil Aviation Flight University of China, Guanghan, China

Correspondence should be addressed to Han Songchen; hansongchen@scu.edu.cn

Received 11 January 2021; Revised 21 March 2021; Accepted 29 March 2021; Published 14 April 2021

Academic Editor: Peng Ji

Copyright © 2021 Tang Zhixing 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.

Flight delay is one of the most challenging threats to operation of air transportation network system. Complex network was introduced into research studies on flight delays due to its low complexity, high flexibility in model building, and accurate explanation about real world. We surveyed recent progress about flight delay which makes extensive use of complex network theory in this paper. We scanned analyses on static network and temporal evolution, together with identification about topologically important nodes/edges. And, we made a clarification about relations among robustness, vulnerability, and resilience in air transportation networks. Then, we investigated studies on causal relations, propagation modellings, and best spreaders identifications in flight delay. Ultimately, future improvements are summarized in fourfold. (1) Under Complex Network, flight operation relevant subsystems or sublayers are discarded by the majority of available network models. Hierarchical modelling approaches may be able to improve this and provide more capable network models for flight delay. (2) Traffic information is the key to narrow the gap between topology and functionality in current situations. Flight schedule and flight plan could be employed to detect flight delay causalities and model flight delay propagations more accurately. Real flight data may be utilized to validate and revise the detection and prediction models. (3) It is of great importance to explore how to predict flight delay propagations and identify best spreaders at a low cost of calculation complexity. This may be achieved by analyzing flight delay in frequency domain instead of time domain. (4) Summation of most critical nodes/edges may not be the most crucial group to network resilience or flight delay propagations. Effective algorithm for most influential sequence is to be developed.

1. Introduction

Commercial flights in China mainland has achieved 1.76 times growth in just one decade, from 4.22 million in 2008 [1] to 11.66 million in 2019 [2]. Except for strong demand to satisfy, air transportation network system may be disturbed by severe weather and airspace restrictions heavily. Most of flight delays are caused by them [2]. Besides, this critical infrastructure must be resilient or robust in case of earthquakes, system failures, and other unexpected situations. Microcomputer simulation is able to model air transportation network system at high resolution and accuracy via professional software [3, 4]. It once was regarded as the most efficient approach for airspace capacity, network

resilience, and flight delay. However, this method did not become the first candidate for national, intercontinental, worldwide, and other large-scale air transport network systems. It is mainly because of its high complexity in model building and strong inflexibility to model structure modification.

It is much appreciated that network science is able to simplify a complex system so that we can better understand its function as a whole [5]. Despite aged less than two decades, complex network theory [6, 7] in network science has experienced a tremendous growth. Many real systems [8–10], composed of a large set of interacting elements, can thus get accurately interpreted. It is not surprising that this methodology was introduced into researches on flight delay.

Studies about network character and network resilience also got stimulated even though they contribute to flight delay phenomenon indirectly.

Network properties root for characterization, reproduction, and even prediction of the network system [11]. With strong fluctuations at various time scale, not only network properties evolved but also flight delays diffused. Complex network was applied to model and analyze Worldwide Airport Network [12], USA Airport Network [13], and China Airport Network [14] in the very beginning of the theory. Metrics from complex network were also utilized to characterize the temporal evolutions of airport networks worldwide [15–20].

Flight delay will occur frequently and propagate heavily if air transportation network is vulnerable, less robust, or less resilient to disturbances. Network efficiency [21] and size of the largest connected subgraph [22] became the most frequent utilized indicators in robustness description of air transportation network [23–28] and both of them come from complex network. Indices, including degree, betweenness, clustering coefficients, and closeness, could be employed to make effective and efficient attack/recover strategy [23–28] to network resilience.

On the foundation of character analyses about flight delay in complex network, scholars attempted to discover causal relations of flight delay [29–34]. Modellings or predictions about flight delay propagation in network level have also achieved much progress [35–40] with support of complex network.

With advancement in studies on network character analysis, network resilience, and flight delay, there emerged a new hot issue: identification of influential vertices or links. The identification is crucial to achieve certain structural or functional objectives. They are topologically important nodes/edges in network character analysis [41–44], influential nodes/edges to network resilience [24, 45], and best spreaders for flight delay [34, 40, 46]. However, the identification is not a trivial task due to challenges in balancing local and global indices and parameter-free and multiparameter indices [47].

This paper tracks and concludes recent progress in network character analysis (Section 3), network resilience (Section 4), and flight delay (Section 5). The survey about identification of influential vertices/links is conducted in these 3 sections since we are concentrated on flight delay under complex network. Section 2 will introduce the basic concepts about air transportation network and complex network. Finally, Section 6 is arranged to summarize the future directions.

2. Basic Concepts

2.1. Air Transportation Networks. Static air transportation network comes from one snapshot of the network states or the accumulation of network states within certain time duration. There are five categories of interesting air transportation networks under complex network: (1) Airport Network, (2) Airline Network and Air Alliance Network, (3) Air Route Network, (4) ATC Sector Network, and (5)

Country Network. The *Airport Network* is built by connecting pairs of airports if they have a direct flight, which are illustrated by dotted lines in Figure 1. Consider that flights may be operated by different airline companies, and there are different types of dotted lines in Figure 1. And, several airline companies may ally to share their flights and to provide more convenient transport choices. If we discriminate the flights with their airline company [48] or with their airline alliance, the *Airline Network* or *Air Alliance Network* gets founded. In some airport networks or airline networks, there are more than one airport in the same city. These airports may be merged into one node. Traffic information outside the city is accumulated into the new node, while traffic information inside the city is neglected [49]. What is noticed is that airline network differs from airport network only in the airline company. If the airline company information does not get counted, the airline network is the same as airport network, for example, [49, 50].

Nevertheless, flights will not fly to its destination airport in a straight airline in real air transportation operation. For safety and airspace capacity, they have to follow the predefined nominal track, named air route or airway. Usually, an air route looks like a polyline rather than a straight line. Air route/airway is consisted of several intermediate waypoints, airports and segments among them, i.e., W1, W2, and W3, and the solid lines among them, as shown in Figure 1. *Air Route Network* is created once these air route waypoints, airports, and predefined nominal tracks get considered. Besides, air traffic controllers are responsible for the flight safety and efficiency within pre-established airspaces, named air traffic control (ATC) sectors. Usually, an ATC sector covers several waypoints of different air routes and zero or more airports, as illustrated by Figure 1. *ATC Sector Network* is constructed by linking ATC sectors if they have an air route segment. In *Country Network*, all the airports, waypoints, or ATC sectors within the same country or region are merged into one vertex, and the connections inside the country or region are discarded. Since air route network and ATC sector network are most relevant to flight operations, studies conducted on them benefit research studies on flight delay most.

The link in above networks can employ traffic information between nodes as weight. Frequent adopted weights include number of flights, passengers, and available seats. Without weight, the link is a binary. Besides, directional traffic between two same vertices can also get modelled separately. In airport network in Figure 1, edge $A \rightarrow C$ may describe the traffic from A to C while the link $C \rightarrow A$ may represent the traffic in opposite direction, and this is the directed network [51].

2.2. Temporal Network. Temporal network consists of a sequence of static networks over multiple time snapshots. In temporal air transportation network, connections, weights, and architecture evolve with time t , as shown in Figure 2.

Figure 2 is an example of temporal airport network, in which node denotes airport. A link is created when there exists a direct flight between two airports. If we take the

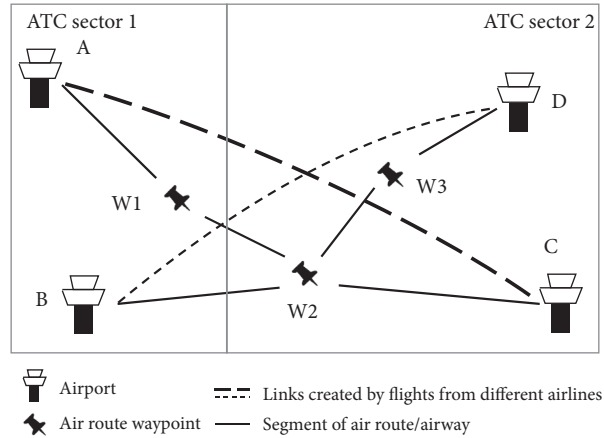


FIGURE 1: Illustration of air transportation network.

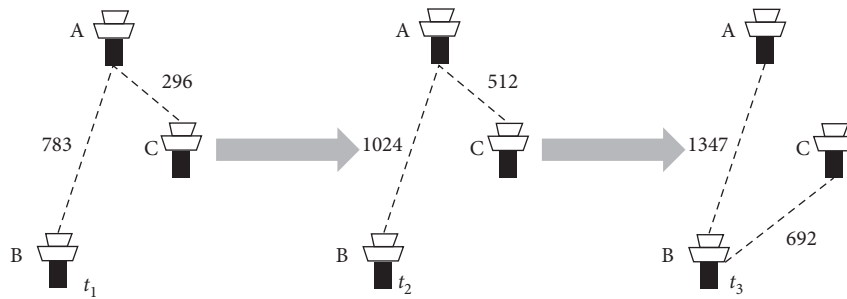


FIGURE 2: Illustration of temporal air transportation network.

number of flights as weight, both weights and connections of airport network may evolve along with time t . Except for structural evolutions, temporal air transport network is also filled with dynamics, such as flight delay.

2.3. Fundamental Metrics in Complex Network. Table 1 lists the most frequent used fundamental metrics in air transportation networks under complex network.

These fundamental metrics are for both undirected and directed networks. And, H-index [56, 57], weighted metrics [58], and other sophisticated methods [17, 20] are developed basing on them. All of these indicators are to describe single vertex or whole network directly. While eigenvectors [59] and k -core [60] are to evaluate nodes via their neighborhoods or layers they belonged to.

3. Network Character Analysis

Network character roots for the characterization, reproduction, and even the prediction of the network system. Modelling and analyzing air transportation network are the fundamental step for research studies on flight delay. This section is to investigate analyses about air transport network characters under complex network on the context of flight delay. Our focuses in this section include static and temporal network, together with the identification of topological important nodes/edges.

3.1. Analysis on Static Network. Static air transportation network comes from one snapshot or the accumulation of network system states within certain time duration. In complex network theory, general procedure to analyze static air transport network is to build the network model basing on collected data, calculate topological metrics of network model, and analyze these indicators.

Traffic information characterizes the connections in air transportation infrastructure and is fundamental for a comprehensive system description. Hence, this section pays special attention to the characterization of traffic information in established network models. For this purpose, researchers usually employ weighted metrics and even develop new indicators. Table 2 compares some strong impact analyses on static air transport networks under complex network theory.

Discussions about Table 2 are in twofold. *The first one* refers to the air transport network model. Despite the topology of different air transport network models being analyzed by worldwide scholars, the airport network model acquires the most attention. It is mainly because it is the least difficult one. However, the airport network model is also the most different one from real air transport operation. It discards plenty of intermediate waypoints and contacts. Thus, factors affecting flights, such as diverge, converge, and cross traffic in the air, airspace restriction, and severe weather, could not get effectively considered. This problem will be further confirmed by demonstrations of Tables 3–6.

TABLE 1: Most frequently used topological metrics in air transport network.

| Metrics | Equation | Interpretation |
|-----------------------------------|---|---|
| Degree [52] | $k_i = \sum_j m_{ij}$ | Where m_{ij} is the connection between node i and node j ; $m_{ij} = 1$ if there is a connection existing; $m_{ij} = 0$ otherwise; this metric refers to the number of connections with other nodes in the network |
| Betweenness [53] | $BC_i = \sum_{j \neq k \neq i} (\sigma_{jk}(i) / \sigma_{jk})$ | Where σ_{jk} is the number of shortest paths going from node j to node k ; $\sigma_{jk}(i)$ is the number of shortest paths going from node j to node k and passing through node i |
| Average shortest path length [52] | $L = (1/(n(n-1))) \sum_{i,j \in N, i \neq j} l_{ij} = (1/n) \sum_{i \in N} l_i$ | Where N is the set of all nodes in the network, n is the number of nodes; l_{ij} [54] is the length of the geodesic from node i to node j , the minimum number of edges connecting from node i to j ; |
| Efficiency [21] | $E = (1/(n(n-1))) \sum_{i,j \in N, i \neq j} (1/l_{ij})$ | $l_i = (\sum_{j \in N, j \neq i} l_{ij}) / (n-1)$ |
| Clustering coefficient [55] | $C_i = (\sum_{j,k \in K_i, j \neq k} m_{ij} m_{ik} m_{jk}) / (k_i(k_i-1))$ | Where K_i is the set of all the neighbor nodes of node i , k_i is the number of nodes K_i ; m_{ik} is the connection between node i and node k ; this metric gives an overall indication of how nodes are embedded in their neighborhoods |

TABLE 2: Some strong impact analyses on static air transport networks.

| Region | Type | Time period | Directed | Weight | Metrics | | Ref | |
|-----------|-----------------|----------------------|----------|---------------------------|------------------------|-----|------|------------------------|
| | | | | | Conventional | New | | |
| Worldwide | Airport network | 2000.11.1- | No | — | Degree | — | [12] | |
| | | 2001.10.31 | | | Betweenness | | | |
| | | | | | Shortest path length | | | |
| Worldwide | Airport network | 2002 | No | Number of available seats | Degree | — | [58] | |
| | | | | | | | | Strength |
| | | | | | | | | Shortest path length |
| Worldwide | Airport network | 2011 | No | Number of flights | Clustering coefficient | — | [50] | |
| | | | | | | | | Degree |
| | | | | | | | | Clustering coefficient |
| USA | Airport network | — | Yes | Number of flights | Degree | — | [13] | |
| | | | | | | | | Shortest path length |
| | | | | | | | | Clustering coefficient |
| USA | Airport network | 06:00–24:00 2010.1.1 | Yes | Number of flights | Degree | — | [61] | |
| | | | | | | | | Shortest path length |
| | | | | | | | | Clustering coefficient |
| Europe | Airport network | 2014.8 | No | Number of flights | Degree | — | [62] | |
| | | | | | | | | Clustering coefficient |
| | | | | | | | | Degree |
| China | Airport network | — | Yes | Number of flights | Shortest path length | — | [14] | |
| | | | | | | | | Clustering coefficient |
| | | | | | | | | Degree |
| China | Airport network | 2015 | No | — | Degree | — | [49] | |
| | | | | | | | | Degree |
| | | | | | | | | Shortest path length |
| India | Airport network | 2004.1.12 | Yes | Number of flights | Clustering coefficient | — | [63] | |
| | | | | | | | | Clustering coefficient |

Another problem about air transport network model lies in the category of nodes. Although Verma et al. [50], Lordan and Sallan [62], and Du et al. [49] encapsulate airport networks into multilayer infrastructures via the “k-core decomposition” method, vertices in all the five air transport network models denote airport, region/country, air route waypoint, or ATC sector alternatively. In essence, nodes in all available models only represent points of flight path or flight route. Airport, air route waypoint, and ATC sector will not attend on the same network model. This simple modelling method may work well in network-level description. However, there are various factors which may affect flight operation, such as plane rotation, flight crew, air traffic controller, and “communication, navigation, and

surveillance facilities.” Any of them may produce large-scale disruptions to air transport networks. Without these flight operation relevant subsystems and sublayers, the current modelling scheme is hard to be effective for flight delay.

The second one pertains to the characterization about traffic of air transport network system. Number of flights and number of available seats are the most frequent employed weights to signalize network traffic in almost all relevant literatures, including those in Tables 2–6. However, a comprehensive description of air transport network system requires much more information, such as airspace capacity, geographical length, direction, and altitude restriction of a segment of air route/airway. And, to the best of our knowledge, they are discarded by current investigations.

TABLE 3: Some strong impact temporal evolution analyses on air transport network.

| Region | Type | Time period | Directed | Weight | Metrics | | Ref |
|-----------|-------------------|-------------|----------|---|---|---|------|
| | | | | | Conventional | New | |
| Worldwide | Country network | 2002–2013 | No | Number of passengers | Degree Betweenness Shortest path length Clustering coefficient | — | [15] |
| USA | Airport network | 1990–2000 | No | Number of passengers Number of flights | Degree | Weight growth lifetime of links | [16] |
| Brazil | Airport network | 1995–2006 | Yes | Number of passengers Amount of cargo carried | Degree Shortest path length Clustering coefficient | Entropy | [17] |
| USA | Airline network | 1990–2007 | No | — | Degree Betweenness Shortest path length Clustering coefficient | — | [18] |
| China | Airport network | 2002–2010 | No | — | Degree shortest path length clustering coefficient | — | [19] |
| China | Air route network | | | | | | |
| Europe | Airport network | 2011–2013 | Yes | Number of available seats | Degree betweenness shortest path length clustering coefficient | Cov (ratio of the standard deviation to the mean) | [20] |
| Europe | Air route network | | | Number of flights | | | |

Moreover, present research studies are unable to employ more than one weight. Considered information is so limited that it is hard to make a comprehensive explanation of the air transport network system.

Besides weights, scholars also employed topological metrics and correlations among them to reveal features of air transport network. Based on those conventional metrics in Table 2, Barrat et al. [58] proposed weighted clustering coefficient and weighted average nearest-neighbors degree. Correlations between indexes, for instance, relations between clustering coefficients-degree [61, 63] and strength-degree [58], get further evaluated. Guimera et al. [65] and Bianconi et al. [66] also investigated the connections between community structure and air transportation network topology. Bianconi et al. [66] proposed an entropy measure-based indicator to quantify the dependence of USA Airport Network structure on community structure and individual vertex's degree.

Much progress has been achieved by the metric-based analyses. However, the majority of available metrics are topological ones, including all the conventional and their weighted indices in Tables 2–6. They are to characterize the air transport network topology instead of the network functionality, while air transport network is a physical network system which is born for traffic. There exists an inherent gap between topology and functionality in current studies and the gap remains unnarrowed.

3.2. Temporal Evolution Analysis. During strong fluctuations at various time scales in air transportation network, not only topological properties evolved but also flight delay propagated. As temporal network consists of a sequence of static networks over multiple time snapshots, analyses about temporal evolutions are mainly conducted through statistics about network topological metrics and their relations in different moments.

Besides the focuses on static network in Section 3.1, primary concern in this section lies in how to analyze the temporal evolutions. Table 3 surveys some strong impact analyses on temporal evolutions.

Table 3 confirms the conclusions in Section 3.1. It also reveals that most investigations in this field are performed via basic statistics on network indices and their relationships.

Wandelt and Sun [15] analyzed the yearly and monthly evolutions of worldwide country network through degree and density (a degree-related indicator). Relationships between degree and betweenness and topological critical nodes and links got further displayed. Besides, they calculated yearly correlations within unweighted degree and passenger-weighted betweenness and correlations within functionally critical nodes and weighted links. Gautreau et al. [16] discovered topological characters in USA airport network were stable via statistics of degree, passenger-weighted degree, and passenger flows. Moreover, they explored the

microscopic dynamic of network through analyses on weight growth and lifetime of links. da Rocha [17] scanned the yearly evolutions of unweighted and weighted Brazil Airport Network via degree, shortest path length, and clustering coefficient, together with their distributions and correlations among them.

Cai et al. [19] analyzed temporal evolution of Chinese Air Route Network via early distribution of flight-weighted degree, yearly traffic flows, and traffic flow growth rates. Based on degree, weighted degree, clustering coefficient, betweenness, weighted betweenness, closeness (shortest path length-based metric), and weighted closeness, Sun et al. [20] studied temporal evolutions of European Air Route Network and European Airport Network. Adopted weights are number of flights and number of available seats.

Temporal tendencies of metrics and their correlations are mainly presented and discussed via time-figures. Principle-level analysis is quite rare. Weight growth and lifetime of links [16], entropy of the degree distribution [17], and variations and CoV of conventional indicators [20] are employed to improve traditional methods and to narrow the gap between topology and functionality of air transport network.

3.3. Identification of Topological Important Nodes/Edges. It was recognized that identification of influential vertices or links [42–44] is crucial for certain benefits. In essence, the identification in this section is to evaluate nodes/edges with indicators in static networks. However, no single index can perform the duty. The general detection idea is based on multitopological attributes, such as AHP [67]. Thus, we are interested in the idea of identification method, metrics they fused, and performance test approaches. Table 4 is organized by the concerns.

Li and Xu [41] employed both functional index and topological metrics in the evaluation of node importance. These indicators are fused by a method based on fuzzy soft set theory which is able to integrate several indices over different time intervals. Performance of the proposed method is evaluated through the change of airport network efficiency after the airport is removed intentionally. Ren et al. [42] proposed a node sorting algorithm based on VCM and four conventional topology metrics. The VCM assumes that the index with larger difference has a large impact on network physical properties and is more important. They adopt SIR (susceptible-infected-recovered) [38] model to obtain node's infection ability. And, performance of VCM and four topology indices are compared through the SIR model. The IEM [43] believes that the indicator with small entropy provides more information and is more crucial.

Depending on topological indices, these approaches [41–43] calculate the importance of a node directly. Instead, Ren et al. [44] measured influence of waypoint through change of network after the node is removed. The change is defined as relative entropy of network agglomeration. Network agglomeration is determined by the average path length which has been introduced in Table 1.

The identification of influential nodes/edges was stimulated greatly by these two kinds of approaches. However, the gap between topology and functionality of the air transport network system remains still unnarrowed. Most frequent utilized indicators are topological ones and the network changes are also quantified via the change of network topology. Last but not least, it is much appreciated that air route network and ATC sector network have aroused attentions from researchers, since they are the most relevant networks to real flight operations.

3.4. Summary. Except for analyses about network characters under complex network on the context of flight delay, we are much concerned about network models which are fundamental to subsequent research studies since air transportation network is a physical network and is born for traffic. Future improvements lie in the bridge between theoretical approaches and air transportation network functionality.

- (1) There are various factors and systems that may disturb a flight, such as weather, plane rotation, flight crew, air traffic controller, “communication, navigation, and surveillance facilities,” diverge, converge, and cross traffic in the air, and airspace restrictions. Any of them may produce large-scale disruptions to air transportation networks. However, as concluded by Section 3.1 and confirmed by other sections, current five network models are unable to take these factors into account due to the absence of intermediate points of flight path and the single category of nodes. This prevents these network models becoming the most effective modelling schemes for flight delay. With flight operation relevant subsystems or sublayer embedded in, such as [68], hierarchical modelling approach may be the most appropriate framework for flight delay.
- (2) Since flight delay is concerned with functionality/operation of the air transport system, the gap between topology and functionality should be taken seriously when aimed at flight delay. More approaches or indices, responsible for both topology and operation of air transport network, are to be developed. Moreover, the comprehensive descriptions and evaluations about air transport network will definitely get stimulated by weights beyond traffic information, such as geographical length, capacity, and altitude restriction, or by network models with more than one kind of weights.

4. Resilience

Flight delay will occur frequently if air transportation network is less robust to disturbances. The word resilience originally originated from the Latin word “resiliere,” which means “bounce back.” Resilience implies the ability of an entity or system to return to its normal condition when

TABLE 4: Comparison of multimetric-based methods.

| Network type | Identification method | Fused metric | Performance/effectiveness test | Ref |
|--------------------|--|--|---|------|
| Airport network | <i>Fuzzy soft set</i> (fuse multi-indices over different time intervals) | Degree, Betweenness Passenger flow | Network efficiency after airport gets removed | [41] |
| ATC sector network | <i>Variation coefficient method</i> (VCM, objective weighting method to aggregate multiattributes) | Degree betweenness closeness eigenvector | SIR model Kendall's tau coefficient | [42] |
| Air route network | <i>Improved entropy weight</i> (IEW, objective weight method to aggregate multiattributes) | — | SIR model Kendall's tau coefficient | [43] |
| Air route network | Through changes in network agglomeration relative entropy when node removed | — | SIR model Kendall's tau coefficient | [44] |

disrupted. Resilience in the engineering system can get clarified via the three phases of system responding in Figure 3 [69].

Performance stands for network system's ability to perform required task. There are various performance indicators in complex networks and this will be discussed later. In the original steady phase, network performance maintains its target level p_0 . In the disruptive phase, system performance drops to the lowest level p_r . In the recover phase, network system performance recovers to new steady level p_{ns} . Inspired by definitions about absorptive capability and restorative capability [70], we argue that, in the air transport network system,

- (1) Robustness refers to network performance loss ($PL = p_0 - p_r$ in Figure 3) when perturbed. Disturbance will produce much less performance loss to a robust network.
- (2) Average drop rate of network performance reflects resistance of network against the perturbation. Sometimes, it can be expressed by the performance loss over specified time, $(PL/\Delta t)$. It can also be measured by the performance loss under certain attack strength $(PL/\Delta a)$. Attack strength a can be quantified by number of removed nodes or edges. Network vulnerability is described via performance loss and its average drop rate.
- (3) $p_{ns} - p_r$ and its average increase rate are utilized to characterize restorative capability of air transportation network. Similar with the average drop rate of PL , $((p_{ns} - p_r)/\Delta t)$ indicates restored performance over specified time. And, $((p_{ns} - p_r)/\Delta r)$ represents restored performance under certain recover strength. Recover strength r can be quantified by number of repaired nodes or edges.

Thus, resilience is composed of vulnerability and restorative capability. Relationships among resilience, vulnerability, and robustness are illustrated in Figure 4.

Following the tagged numbers in Figures 3 and 4, relations among robustness, vulnerability, and resilience are easy to be clarified. And, this clarification breeds a clear sight into numerous research studies about resilience of the air

transportation network system. Current investigations can be reviewed based on the following 4 questions.

- (1) How much is the network performance loss when it is interrupted? How to measure this robustness?
- (2) What is the most effective and efficient attack strategy or what can be done to maximize the performance loss? Whose removal decreases network performance most?
- (3) What is the best strategy to recover the network system's performance? Whose recovery increases network performance most?
- (4) How to make a comprehensive description about the resilience of the air transportation network system?

4.1. Robustness Measurement. In early studies about error and tolerance in complex network, Albert et al. [71] regarded average length of the shortest paths between any two nodes in the network as the performance of complex network. Motter and Lai [22] quantified network performance in terms of the relative node size of the largest connected subgraph.

Our concerns in this section are about question (1), performance indicator, and how to measure the robustness of air transport network. Table 5 presents recent impact literatures about robustness. These studies are to remove or recover vertex and its connected edges one by one based on different select criteria. The network robustness is measured and analyzed by the adopted performance indicator.

Table 5 reveals that network efficiency [21] along with size of the largest connected subgraph ranks the most frequently utilized performance indicators of air transportation network. The network efficiency closely relates to average length of shortest path as illustrated in Table 1. Both network efficiency and size of the largest connected subgraph are to signalize network performance from perspective of network connectivity other than network functionality. Nevertheless, air transportation network is practical infrastructure which is born for air traffic. Network robustness is concerned with system functionality, as discussed previously. Without

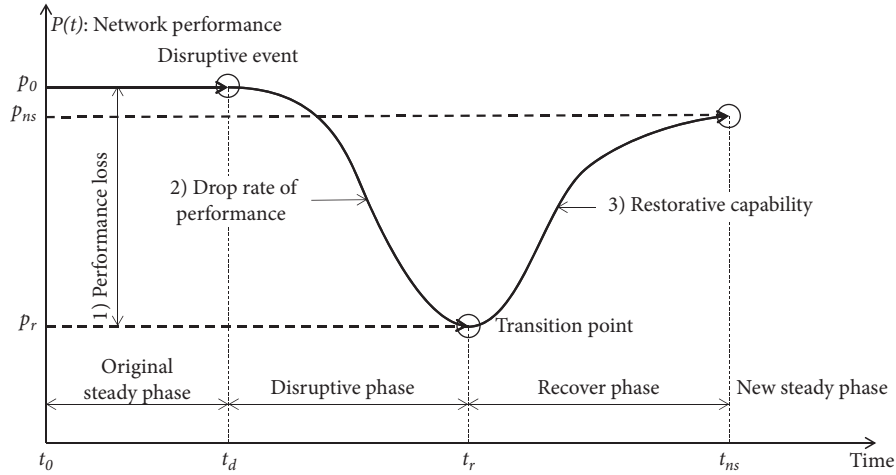


FIGURE 3: Network resilience profile.

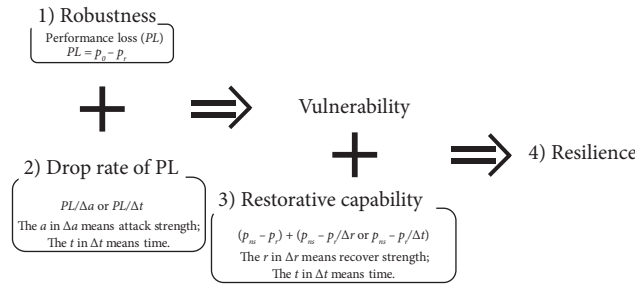


FIGURE 4: Illustration of relationships among resilience, vulnerability, and robustness.

TABLE 5: Recent impact literatures about robustness in air transport network.

| Region | Network type | Indicator of network performance | Node select criteria | Ref |
|-----------|--|--|---|------|
| Worldwide | Static airport network | Size of the largest connected subgraph | Degree, betweenness, and other 3 topological metrics | [23] |
| Worldwide | Static airport network | Size of the largest connected subgraph | Degree, betweenness, and 2 topological metrics | [24] |
| USA | Static airport network | Size of the largest connected subgraph | Degree-weighted degree | [25] |
| Canada | Static airport network | Size of the largest connected subgraph | Weighted clustering coefficients betweenness connectivity (weighted degree) closeness | [26] |
| German | Static airport network | Size of the largest connected subgraph network efficiency survived links | Degree, betweenness, eigenvector closeness, and other 2 topological metrics | [64] |
| China | Static airport network static air route network static ATC sector work | Size of the largest connected subgraph network efficiency | Degree betweenness closeness | [27] |
| China | Static ATC sector network | Size of the largest connected subgraph network efficiency | Degree-weighted degree betweenness | [28] |

fundamental traffic information, these two topological metrics are less convincing to measure robustness of air transportation network system.

On the context of air transport network functionality, Janić [72] relied on number of flights/passengers to calculate robustness of existing airports. And, robustness of removed

airports is measured by their weights. Then, he regarded summed robustness of all left vertices as network robustness. Wandelt et al. [73] and Sun et al. [64] employed the unaffected passengers with rerouting as a baseline index in evaluation of air transportation system robustness. Besides, Sun et al. [64] and Dunn and Wilkinso [74] defined survived

links as robustness of air transportation network. All these summed weight metrics [64, 72, 73] assessed robustness of air transportation network from the aspect of traffic demand.

Another kind of solution comes from the perspective of traffic supply and is based on airspace capacity [75–77]. Pien et al. [75, 76] proposed a new robustness index called the relative area index (RAI). The index quantifies the influence of individual node to the performance of the entire network when it suffers capacity reduction at a local scale. And, the model estimates the maximum flow of network as its new capacity when the air transport network is influenced. They [75] also conducted a comparative analysis between RAI-based robustness and betweenness-based robustness. Yoo and Yeo [77] regarded the redundant capability to replace the damaged node as the adaptive capacity. Furthermore, they evaluated the air transportation network robustness based on adaptive capacity concept.

Much progress has been achieved to describe air transport network performance from perspective of network functionality. They either rely on traffic demand or depend on traffic supply. Thus, disadvantages of topological character-based methods get surmounted and gap between functionality and topology of air transport network get narrowed. Nevertheless, few take traffic demand and traffic supply into their description simultaneously. This goes against with the indication of COVID-19. In the rare scenario, traffic demand is suppressed heavily, while traffic supply (capacity of air transport network and airspace) remains almost the same as before. The demand is so weak that any random errors or intentional attacks to network system would produce much less disturbances than before. Traffic demand and traffic supply are suggested to get considered simultaneously in the measurement of air transport network robustness.

4.2. Attack and Recover. This section is to discuss questions (2) and (3) in the beginning of Section 4, search most effective and efficient strategy to reduce or to restore network performance, and locate those nodes or edges influential to network performance. Table 5 reveals that most adopted strategy is to remove or recover nodes one by one basing on their values of degree, betweenness, clustering coefficients, closeness, or other topological metrics. And, strategy effectiveness is evaluated via the estimation of network robustness. However, these nodes' select criteria are also topological ones since they completely rely on topological indices and robustness indicators. In the birth of this functionality-oriented research, there were no other better solutions, and researchers had to go on with topological methods. However, much progress has been achieved now. More effective and efficient strategies should have been developed.

Attack Strategy. Inspired by general game playing process, Wandelt et al. [73] proposed a new exploration/exploitation search technique to find attacking strategies efficiently. It is based on a Monte Carlo Tree Search algorithm, rather than on network metrics. Dunn and Wilkinson [74] assumed that there may be

unused airport capacity to accept previously unscheduled flights. This permits an adaptive saving and rerouting edges. This adaptive strategy is demonstrated to be effective in raising robustness. Thompson and Tran [78] utilized a three-stage optimization model (defender-attacker-defender) to analyze USA air transportation network robustness. The model approximated the dynamics among three opposing agents: an operator that seeks to minimize the network's operational cost via optimal passenger rerouting, an attacker that aims to maximize that cost by disrupting a number of network routes, and a defender that is tasked with mitigating the actions of the attacker by protecting key routes.

Recover Strategy. Since recover strategy in air transportation network system has not aroused much interest, current recover strategies mostly rely on topological characters. Clark et al. [25] evaluated the performance of three recover strategies in USA airport network. They are based on traffic volume and random and network centrality (eigenvector, closeness, betweenness, and degree), respectively. Wang et al. [28] compared the effects of random, degree-based, betweenness-based, and remove sequence-based recover strategy in China ATC sector network. However, to the best of our knowledge, we have not noticed much progress in recover strategies. Current relevant literatures neither take traffic information into account nor study recover strategies through the modelling of network system evolutions such as [73, 74, 78].

Influential nodes or edges to network robustness can be identified by the attack and recover strategies. However, these diagnosed nodes/edges are independent individuals and summation of them may not be the most crucial set. Du et al. [45] employed a memetic algorithm to seek the minimum network robustness after removing certain edges in Chinese air route network. The attacking performance of the algorithm is superior to the highest edge-betweenness adaptive strategy. And, the solution of the algorithm is the vital set of edges. Soria et al. [24] proposed two customized methods to search most effective sequence of edges in network robustness reduction. It is based on betweenness and damage, respectively. They also compared the performance of genetic algorithm, simulated annealing, two customized algorithms, and a combination of the two proposed methods in worldwide airport network. Simulation experiments demonstrated that the combined one is the most effective.

4.3. Resilience Description. It is quite rare to meet literatures which are answering question (4). Xu [27] and Wang et al. [79] defined a general resilience index for air transportation network in which absorptive and restorative capability (illustrated in Figure 3) get counted. Through the evaluation of system performance, comprehensive resilience of China Airport Network was analyzed.

It is much treasured that this integration of robustness, resistance to attack, and restorative capability covers all response phases to perturbations. However, the air transportation network is a practical/physical system, such as the electric power supply system [70]. Its operation is filled with interactions among subsystems or sublayers, *i.e.*, airport, “communication, navigation, and surveillance facilities,” air traffic controller, etc. Operation of subsystem or sublayer is so sophisticated that any study on subsystem or sublayer is worthwhile. These complicated subsystems or sublayers should get further accounted in resilience descriptions.

4.4. Summary. We clarified the relations among robustness, vulnerability, and resilience in the beginning of this section. Under this guidance, we surveyed literatures about resilience of air transportation network under complex network. Future efforts are outlined as follows:

- (1) Air transportation system is a practical infrastructure, whose robustness should be connected with system performance. Traffic information is fundamental for robustness description. Except for topological indicators, scholars have developed numerous indexes to characterize air transportation network robustness from perspective of either traffic demand or traffic supply. However, the situation in COVID-19 suggests that traffic demand and traffic supply should get counted simultaneously in the measurement about air transport network robustness. In the rare scenario, traffic demand is suppressed heavily, while traffic supply (capacity of air transport network and airspace) remains almost the same as before. The demand is so weak that any random errors or intentional attacks to the system would produce much less disturbances than before.
- (2) Under complex network framework, conventional attack strategy and recover strategy are to remove and recover nodes/edges one by one based on the values of topological metrics. Monte Carlo Tree Search algorithm [73], adaptive saving and rerouting edge strategy [74], and defender-attacker-defender optimization model [78] have been employed for more effective attack strategy. While, to the best of our knowledge, investigations about recover strategy still rely on topological indices.
- (3) It is pretty appreciated that resilience description about air transportation network system has proceeded to the cover of all response phases, although it just began. There are various subsystems or sublayers in air transport networks and any of them may produce large-scale disruptions. These factors should get further considered in resilience descriptions. We believe that only in this way resilience of the air transport network system can get accurately evaluated.

5. Flight Delay

Flight delay occurs when airplane cannot takeoff or land on time. This abnormality may spread to downstream flight through airplane rotation. Furthermore, flight delay will propagate heavily in some extreme situations, such as severe weather and airspace restrictions. Under complex network framework, flight delay diffusion can be portrayed as network dynamics, such as cascade failure [22, 80] or epidemic process [81].

Current investigations about flight delay in complex network theory focus on empirical analysis, reveal of delay causal relations, prediction or modelling of delay propagation, and identification of best spreaders. Empirical analyses [27, 82] share similar assumptions, approaches, and limitations with network character analysis in Section 3. Therefore, this section will not discuss them anymore. And, these articles [83–87] are also off our concentrations since they are beyond complex network framework, even though they are worthwhile and well qualified in the prediction or modelling of delay propagation.

5.1. Flight Delay Causal Relations. On the foundation of character analyses about flight delay in complex network, scholars attempted to discover causal relations of flight delay. The causal relations among vertices may get mined via significance test methods in statistics. Flight delay causality network is built by connecting pairs of vertices if they have a direct induced flight delay. Moreover, characters of flight delay network are analyzed under complex network theory.

Table 6 surveys recent impact studies on flight delay causal relations under complex network framework. In current situation, properties of established flight delay networks are evaluated via topological metrics in complex network. Network metrics-based analyses in this field share similar limitations with network character evaluation in Section 3. Therefore, our primary concern is the fundamental data, especially network work type and time series of flight delay for significance test. And, approaches to mine causal relations among flight delays are also interesting.

Analyses about Table 6 are organized according to the focuses. *The first one* refers to the fundamental data. Despite flight delay causality under complex network theory emerged only six years ago, related studies have covered major nations/regions around the world. The significance test method cannot work without time series of flight delay. Data in time series differ from landing delay to departure delay. Real flight operation prefers ground delay strategy than air delay strategy for safety and air fuel saving. This makes delay often occur in takeoff rather than in landing. Cancelled flights should get counted since considerable flights may be cancelled in large flight delay events.

However, all flight delay causal relations are studied in airport network. Nevertheless, it is air route network or ATC sector network that real flight operates in. There are multiple

TABLE 6: Recent impact studies on flight delay causality.

| Region | Time period | Network type | Data in time series | Significance test method | Network property metric | Ref. |
|--------------|----------------------|-----------------|---|---|---|------|
| Europe | 2011.3.1–2011.8.9 | Airport network | Average landing delay within one hour | Granger causality test extreme events identification-based test | Link number Efficiency Largest shortest path length Transitivity Diameter Assortativity Link number Efficiency | [29] |
| China | 2015.8.1–2015.8.31 | Airport network | Average landing delay within one hour | Granger causality test | Largest shortest path length Transitivity Assortativity Degree | [30] |
| China | 2012.7.1–2012.7.31 | Airport network | Average departure delay within one hour (with consideration of cancelled flights) | Granger causality test | Average cluster coefficient Largest connected cluster Reciprocity parameter Community Motifs Degree Link number Efficiency | [31] |
| China | 2016.12.1–2016.12.31 | Airport network | Average landing delay within one hour | Low-dimensional approximation of condition mutual information for transfer entropy test | Average cluster coefficient Assortativity Motif Modularity Betweenness Diameter | [32] |
| USA China | 2012.8.1–2013.8.31 | Airport network | Average landing delay within one hour | Pearson correlation test | Average path length Average cluster coefficient Link number Average path length | [33] |
| USA | 2015.1.1–2015.3.31 | Airport network | Total departure delay within 15 minutes | Granger causality in tail test | Average path length Average cluster coefficient | [34] |

intermediate nodes between origination airport and destination airport in real flight operation, as illustrated by Figure 1. Flights will frequently diverge, converge, and cross in intermediate nodes, such as W2 for A-C and B-D in Figure 1. Flight delay will diffuse heavily once the intermediate node is blocked or covered by severe weather. Another frequent situation is that once there are too many flights between A-C, W2 might be allocated to these flights with priority. Thus, flights between B-D will be delayed despite there being no direct links between A-C and B-D. In this case, flight delay should be studied in air route network or ATC sector network.

The second concern is about the significance test method adopted to detect flight delay causal relations. The most frequent employed approach is Granger Causality test

which is linear due to the adoption of smoothed and averaged data. However, flight delay usually propagates in a nonlinear fashion. In large flight delays, flight delay might diffuse heavily in one specific time window. However, these heavy propagations are likely to be ignored by the smoothed and averaged data in the classical Granger causality test. In order to overcome these limitations, Belkoura and Zanin [29] developed a causality test method for extreme events to identify those higher than expected delays. It relies on the abnormal values and the statistics of how such values are propagated. They also compared performance of new approach with conventional Granger causality through causal links, transitivity, efficiency, largest shortest path length, greatest distance between any pair of vertices, and assortativity.

Except for averaged data, weighting small and large delays equally also makes Granger causality unable to detect large delays. Small delay is easily absorbed by the flight trip in air and flight schedule buffers [34]. For this reason, Mazzarisi et al. [34] proposed an extension use of the Granger causality test, namely, Granger causality in tail. It only considers extreme events and large delays. They also compared performance of the proposed method with the traditional Granger causality test through causal links, average path length, and average clustering coefficient.

Besides the Granger causality test, Xiao et al. [32] proposed low-dimensional approximation of condition mutual information for transfer entropy test to mine causality among nonlinear flight delays. Then, they designed a simulation experiment based on artificial nonlinear time series. Via true positive-, false negative-, true negative-, and false positive-based indicators, performance of the proposed method is analogically analyzed with Granger causality and transfer entropy approaches. Simulations demonstrated that the estimation accuracy increased when transfer entropy was used to quantify and validate delay propagation. Wang et al. [33] applied Pearson correlation coefficient to capture the correlations between flight delays in different airports. They also established flight delay networks from operational data.

Detection about delay propagation is fundamental for subsequent flight delay researches. Under complex network theory, there emerged Granger causality test, transfer entropy test, and Pearson correlation test methods in current flight delay causal relation studies. Scholars have realized their limitations and attempted to improve them. However, this problem has not gained much attentions in general, and we have not noticed much efforts. Furthermore, evaluations about detection accuracy deserve more attentions. Belkoura and Zanin [29] and Mazzarisi et al. [34] compared accuracy of proposed methods with the traditional method. Xiao et al. [32] evaluated accuracy of the proposed scheme through flight delay propagation simulation experiment. No one has been testified, validated, or evaluated by real flight delay data. More effective flight delay propagation test methods with high accuracy are to be developed.

Ultimately, the idea of above flight delay causality detection methods is to mine delay causality via significance test methods in statistics and the flight delay time series. In essence, these kinds of schemes totally rely on flight delay data to carry out their duty. They do not make full use of flight schedule and flight plan. In real flight operation, there are flight schedules to arrange flights. Most aircrafts are scheduled with several flights within one day. And, every commercial flight has to submit its flight plan in advance. Flight plan contains origination airport, destination airport, planned air route, expected departure/arrival time, and tail number of aircraft. Through tail number, we are able to track flight path of an aircraft and of course the delay propagation path. We believe that accuracy of flight delay causality detection mechanisms will increase once flight schedule and flight plan are accounted in.

5.2. Flight Delay Propagation Modelling. Subsequently, scholars commenced modelling or predicting flight delay diffusions for delay mitigation and reduction. Under complex network framework, flight delay diffusion can be portrayed as network dynamics, such as cascade failure [22, 80] or epidemic process [81].

Our primary concern is about how to model the spreads of flight delay. Most frequent utilized models are SIR [38] in epidemic and cascade failure [22, 80] models. The SIR model divides all the individuals in the system into several groups with different states. Everyone is aligned with only one status at one time step. Delayed flights or delayed airports can be in the infected state, those recovered from flight delay are in the recovered state, and undelayed are in the susceptible state. Flight delay propagation is simulated through state transmissions.

Baspinar and Koyuncu [35] and Mou et al. [36] assumed that recovered vertices are not immune to flight delay and can be delayed with the same possibility with susceptible ones. They employed the simplest SIR model. There are only two kinds of states: susceptible and infected. Before infected nodes recover to normal/susceptible, it will maintain infected status and spread flight delay to neighbors constantly. The SIR model directly labels a node as delayed or undelayed. In real flight operation, airport, waypoint, ATC sector, or other components contains more than one flight. Hence, the accurate transmission rate (infection rate or recovery rate) is of great importance for the predicting or modelling since not all flights in one airport seem to be delayed or undelayed in any moments.

The cascade failure model [22] reproduces flight delay propagations based on flight flow movement in the air transport network system. It assumes that failure or flight delay occurs once the load of node is over than its capacity. Failed nodes have to be removed from the network and their loads have to be reallocated to other connected nodes in air transport network. This reallocation may produce cascade failures. Thus, flight delay diffusions in air transport network get simulated. Wu et al. [37] assumed that delayed airports were able to perform certain amount of their traffic load and kept delayed airports other than closed them. This idea is similar with real flight operation and makes the proposed model more reasonable. They also explored the association of cascade failure [22] and SIR. Allocation strategy of cascade failure remains a challenging issue since there is more than one flight in a single vertex.

In addition, the executions of both individual-based SIR [35, 36] and improved cascade failure [37] require basic computer simulation. This requirement not only increased calculation complexity but also decreased flexibility to the adjustment of network architecture. Nevertheless, scientists are able to draw network system characters and its dynamic process via mathematical equations in the traditional nonindividual-based SIR model. This kind of methods assumes that all the individuals are equally contacted with each other and network is fully mixed. With graph Laplacian

operator [39], Estrada [40] discovered and proved characters of the propagation process. He obtained the rate of convergence and best spreaders through theoretical analysis. In this way, flight delay prevailing can be modelled at a low cost of modelling complexity.

Second concern is about the infection/recovery rate and the reallocation strategy. It is of great significance to resolve them since they are the key parameters to model flight delay propagations.

Baspinar and Koyuncu [35] estimated the infection rate for airport-based SIR model through the data-driven statistical analysis on flight plan information and actual flight flow data. Ground-waiting time of flight is utilized to determine infection rate in the flight-based model. Moreover, they calculated theoretical recovery rates of the airport-based model and flight-based model via equations in the SIR framework. Flight time and ground-waiting time are also considered in the running of simulation. Furthermore, the infection rate and recovery rate from actual flight flow data was extracted by the Euler Method. The accuracy of established models in capturing delay spreading was demonstrated via the comparison between estimated rates and real rates. Mou et al. [36] introduced two airport-based SI models to evaluate effects of flight time and ground-waiting time on the spreads of flight delay. In one of them, infection of flight delay completed at the moment flight lands. In the other one, infection of flight delay accomplished once flight is airborne. Infection rates in both models are directly defined as 0.5, while recovery rates are 0.

Wu et al. [37] did not define or calculate the infection rate and recovery rate directly. They reallocated the exceeded load of infected airport to its connected susceptible airports according on their weights. There are multiple available reallocate strategies, such as average, degree-based, and priority-based. As introduced in Section 5.1, we are able to track the flight path of aircraft through flight schedule and flight plan. The flight path also enables us to follow the spread of flight delay in air transport network. However, flight schedule and flight plan should have been extensively explored to obtain more accurate infection rate, recovery rate, and reallocate strategy.

From flight delay propagation modelling approach to key parameters solving, considerable efforts have been made to complete the framework for flight delay propagation under complex network. The pursuit for high accuracy is a fundamental objective. Generally, the accuracy of flight delay propagation models remains unvalidated by real flight operation data. The accuracy test by real flight data also implies a new idea to predict the spreads. Since current SIR and cascade failure models totally rely on prebuild model to simulate or predict flight delay propagations in network, real flight data may be utilized to revise the modelling or prediction. Based on the difference between predictions of

prebuild models and real flight delays, we believe the revision will increase accuracy of the prediction or modelling.

5.3. Best Spreaders in Flight Delay. With the progress in delay propagation modelling, there emerged a new hot issue: identification of best spreaders, which is indispensable for delay mitigation and reduction. The general method is to measure the flight delays once the node or edge is removed one by one. More severe the flight delay is, more critical the removed nodes or edges are.

Besides the strategy employed to identify best spreaders, *our primary concern* is about how to measure the influence of flight delay once nodes or links are removed. Cardillo et al. [68] simulated the behavior of passenger rescheduling when their flights are cancelled or links between airports are removed. Through the quantification of rescheduled or delayed passengers, they explored the effects of link deletion in the European Airline Network. The number of rescheduled or delayed passengers was also utilized to indicate the network robustness. Voltes-Dorta et al. [46] proposed an algorithm to quantify the disrupted passengers due to closure of a given airport. The approach also reallocated the disrupted passengers in alternative travel itineraries based on shortest path length. The proportion of reallocated passengers was employed to measure robustness of European Airport Network. Accumulated delay experienced by the disrupted passengers was to identify the critical airports.

Mazzarisi et al. [34] found that existing centrality metrics neither respect the flight schedule nor consider airline companies that each flight belongs to. Hence, these indices were not capable to characterize the effect of delays in the air transportation network. Then, they employed a new centrality to identify critical airports. It is different from conventional approaches since importance of airports is measured through their influence to flight delay.

It is much appreciated that flight delay is quantified via operational influences other than topological changes that are adopted in the robustness indicator. However, above works run into the same weak point with the first concern of Section 4.2. They identify influential nodes or links through computer simulation and this raises the calculation complexity. Instead, Estrada [40] is able to locate best spreaders through theoretical analysis with graph Laplacian operator [39]. It achieved a low cost of computation complexity.

Our second concern lies in the strategy employed to identify best spreaders of flight delay. To the best of our knowledge, nodes and edges are selected by random or by values of topological characters in current best spreaders identification studies. These identified nodes/edges may be most critical individuals to flight delay, but summation of them may not be the most influential association to flight

delay in the network level. This conclusion is the same with the attack and recover strategy of network resilience in Section 4.2.

5.4. Summary. On the foundation of character analyses about flight delay in complex network, scholars have made great efforts to detect flight delay causalities, to model or predict delay propagations, and identify influential nodes/edges to flight delay under complex network. Section 5 surveyed recent progress about these challenging and pragmatic concerns. Outlook for future is concluded as follows:

- (1) Under complex network framework, almost all the current research studies in Section 5 are conducted in airport network. Nevertheless, it is air route network or ATC sector network rather than airport network that real flight operates in. In these two networks, there are multiple intermediate nodes between origination airport and destination airport. Airport network ignores these multiple intermediate waypoints and prevents us from tracing some frequent flight delays. Investigations in this section should pay more attention to air route network or ATC sector network for a better accuracy.
- (2) Current flight delay causality detection models totally rely on flight delay time series to mine causal relations among nodes. And, few take fully advantage of flight schedule and flight plan especially in the estimations about the infection rate and recovery rate of SIR and in the reallocation strategy of cascade failure. However, flight path of an aircraft can be traced from flight schedule and flight plan. With guidance from flight path, we are able to follow both the delay spreading of a single aircraft and delay propagations in the whole network. Current studies about flight delay under complex network, including flight delay causality detections and estimations about infection rate/recovery rate/reallocation strategy, should explore benefits of flight schedule and flight plan extensively.
- (3) Most investigations on flight delay propagation modelling and best spreaders identification require computer simulations. This not only raises the calculation complexity but also reduces the flexibility to the adjustment of network architecture. With graph Laplacian operator [39], Estrada [40] revealed and proved characters of propagation process through theoretical analysis. They initialed a new course to seek or develop approaches at low calculation complexity. It further reminds that Laplacian transform, Z-transform, Fourier transform, or wavelet transform in signal processing area may be adopted for that purpose. With transformed into frequency domain from time domain, we can perform analyses on flight delay at low calculation complexity.
- (4) Much efforts have been made to reveal flight delay causalities and model flight delay propagations. However, accuracy of causality detections and diffusion modellings have not been validated by real flight data. Accuracy test via real flight data also implies a new idea to predict the spreads. Since current SIR and cascade failure models totally rely on prebuild model to simulate or predict the propagations, real flight data may get utilized to revise modelling or prediction. Once difference between predictions of prebuild models and real flight delays is employed to revise the next time-step predictions, prediction accuracy of flight delay diffusion in the next time step is believed to increase.
- (5) The general strategy to identify best spreaders of flight delay is to measure the flight delay after the nodes or edges are removed one by one. More severe the flight delay is, more important the node or edge is. In this situation, identified nodes/edges may be the most influential individuals to flight delay. However, the reactions among diagnosed individuals are ignored and summation of them may not be the most critical group to the whole network. For the sake of flight delay mitigation and reduction, the most crucial association of nodes/edges is more valuable. More effective algorithms should be applied into the search for most crucial group of nodes/edges to flight delay.

6. Conclusion

Flight delay is one of most benefiting concerns in air transportation system. From the perspective of complex network, we reviewed research studies on network character analysis, network resilience, and flight delay. All of them directly or indirectly contribute to the flight delay phenomenon in the air transport network system. We have mapped the state of art to identify promising approaches as well as to make their limitations and assumptions clear. With detailed conclusions organized in the main context, future directions are summarized in fourfold:

- (1) In the available network models for air transportation under complex network framework, all the nodes or edges are modelled with the same property. These models disregard numerous subsystems and factors behind flight operation. This modelling method prevents us from tracing and analyzing frequent flight delays. Furthermore, real flight operation may be affected by plane rotation, flight crew, air traffic controller, "communication, navigation, and surveillance facilities," and other subsystems or sublayers. It is impossible to describe vulnerability and resilience of air transport network system comprehensively since these factors are outside current complex network framework. It is necessary to utilize hierarchical modelling approaches to take flight operation relevant subsystems, air route

waypoints, and ATC sectors into network models. Hierarchical modelling approaches will make network models in complex network framework more capable for flight delays.

- (2) Flight delay is a matter of functionality or operation of the air transportation system. In current studies on flight delay under complex network, there exists an obvious gap between topology and functionality. Most available network metrics and robustness indicators are topological ones. Transmission rates about flight delay propagations are mainly quantified via topological connections. Traffic information is the key to narrow the gap. More indices with traffic information and responsible for air transport network operations are to be developed. And, network system robustness measurement should take traffic demand and traffic supply into consideration simultaneously, which is indicated by the situation in COVID-19. Furthermore, flight schedule and flight plan are indispensable in the tracing of aircraft flight path. The flight path is valuable for flight delay causality detections and estimations of flight delay transmission rates. Besides, real flight operation data can be utilized for the accuracy test on flight delay causality detections and flight delay diffusion models. We also believe real flight data can be employed to revise the modelling or prediction about flight delay propagations. Accuracy of flight delay propagation models may thus get raised.
- (3) Most investigations about flight delay propagation modellings and best spreaders' identifications require computer simulations. This requirement not only increases calculation complexity but also decreases the flexibility to network architecture adjustment. With graph Laplacian operator [39], Estrada [40] initiated a new course to seek or develop approaches at low calculation complexity through theoretical analysis. It further reminds that Laplacian transform, Z-transform, Fourier transform, or wavelet transform in signal processing may be applied for that purpose. By transforming into frequency domain from time domain, we can perform analyses on flight delay at low calculation complexity.
- (4) In research studies on attack/recover strategies, as well as investigations on identifications of influential nodes/edges to network resilience and flight delay propagations, most frequent adopted select criteria are to choose node/edge by random or by values of topological metrics. Sophisticated algorithms, such as Monte Carlo Tree Search algorithm [73], adaptive saving and rerouting edge strategy [74], and defender-attacker-defender optimization model [78], have been applied for more effective attack strategies. While to the best of our knowledge, recover strategies still rely on topological indices. Furthermore, those identified nodes/edges may be the most influential individuals to network resilience or flight

delay propagations. However, most of current studies ignored the reactions among them and the summation of them may not be the most critical association. Similar to memetic algorithm [45] and custom methods [24], effective algorithms should be developed to search most influential sequence of nodes/edges to network resilience or flight delay propagations.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by National Natural and Scientific Foundation of China under Grant nos. 71573184, 71874081, U1733105, and U1733203, Special Foundation for Science and Technology Development of Sichuan Province from Chinese Central Government under Grant no. 2020ZYD094, Science and Technology Foundation of Sichuan Province under Grant nos. 2020YFS0541 and 2021YFS0319, and Scientific Research Foundation of CAFUC under Grant no. QJ2021-160.

References

- [1] Civil Aviation Administration of China, 2015, http://www.caac.gov.cn/XXGK/XXGK/TJSJ/201511/t20151102_8717.html.
- [2] Civil Aviation Administration of China, 2020, http://www.caac.gov.cn/XXGK/XXGK/TJSJ/202006/t20200605_202977.html.
- [3] Total Airspace and Airport Modeler, <https://ww2.jepesen.com/airspace-solutions/total-airspace-and-airport-modeler/>.
- [4] Airtopsoft, <https://airtopsoft.com/>.
- [5] P. Holme, "Modern temporal network theory: a colloquium," *European Physical Journal B*, vol. 88, no. 9, 2015.
- [6] R. Albert, H. Jeong, and A. L. Barabasi, "Internet—diameter of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 130–131, 1999.
- [7] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [8] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, "The web of human sexual contacts," *Nature*, vol. 411, no. 6840, pp. 907–908, 2001.
- [9] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.
- [10] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [11] A. Barrat, M. Barthélemy, and A. Vespignani, "Introduction to dynamical processes: theory and simulation," in *Dynamical processes on Complex Networks*, pp. 77–91, Cambridge University Press, Cambridge, UK, 2008.
- [12] R. Guimerá and L. A. N. Amaral, "Modeling the world-wide airport network," *The European Physical Journal B—Condensed Matter*, vol. 38, no. 2, pp. 381–385, 2004.
- [13] L. P. Chi, R. Wang, H. Su et al., "Structural properties of US flight network," *Chinese Physics Letters*, vol. 20, no. 8, pp. 1393–1396, 2003.

- [14] W. Li and X. Cai, "Statistical analysis of airport network of China," *Physical Review E*, vol. 69, no. 4, Article ID 046106, 2004.
- [15] S. Wandelt and X. Sun, "Evolution of the international air transportation country network from 2002 to 2013," *Transportation Research Part E: Logistics and Transportation Review*, vol. 82, pp. 55–78, 2015.
- [16] A. Gautreau, A. Barrat, and M. Barthelemy, "Microdynamics in stationary complex networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 22, pp. 8847–8852, 2009.
- [17] L. E. C. da Rocha, "Structural evolution of the Brazilian airport network," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 4, 2009.
- [18] G. Hua, Y. Sun, and D. Haughton, "Network analysis of US air transportation network," in *Data Mining for Social Network Data*, pp. 75–89, Springer US, Boston, MA, USA, 2010.
- [19] K.-Q. Cai, J. Zhang, W.-B. Du, and X.-B. Cao, "Analysis of the Chinese air route network as a complex network," *Chinese Physics B*, vol. 21, no. 2, Article ID 028903, 2012.
- [20] X. Sun, S. Wandelt, and F. Linke, "Temporal evolution analysis of the European air transportation system: air navigation route network and airport network," *Transportmetrica B: Transport Dynamics*, vol. 3, no. 2, pp. 153–168, 2014.
- [21] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, Article ID 198701, 2001.
- [22] A. E. Motter and Y. C. Lai, "Cascade-based attacks on complex networks," *Physical Review E*, vol. 66, no. 6, Article ID 065102, 2002.
- [23] O. Lordan, J. M. Sallan, P. Simo, and D. Gonzalez-Prieto, "Robustness of the air transport network," *Transportation Research Part E: Logistics and Transportation Review*, vol. 68, pp. 155–163, 2014.
- [24] M. Soria, O. Lordan, and J. M. Sallan, "Heuristics of node selection criteria to assess robustness of world airport network," *Chinese Journal of Aeronautics*, vol. 30, no. 4, pp. 1473–1480, 2017.
- [25] K. L. Clark, U. Bhatia, E. A. Kodra, and A. R. Ganguly, "Resilience of the U.S. National airspace system airport network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 12, pp. 3785–3794, 2018.
- [26] Y. Yassien, *Air transportation infrastructure robustness assessment for proactive systemic risk management*, M.S. thesis, Department of Civil Engineering, McMaster University, Hamilton, Canada, 2020.
- [27] X. Xu, *Research on robustness of air transportation network and flight delay correlation*, M.S. thesis, College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2018.
- [28] X. Wang, S. Miao, and J. Tang, "Vulnerability and resilience analysis of the air traffic control sector network in China," *Sustainability*, vol. 12, no. 9, Article ID 3749, 2020.
- [29] S. Belkoura and M. Zanin, "Phase changes in delay propagation networks," in *Proceedings of the 7th International Conference on Research in Air Transportation*, Philadelphia, PA, USA, June 2016.
- [30] M. Zanin, S. Belkoura, and Y. Zhu, "Network analysis of Chinese air transport delay propagation," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 491–499, 2017.
- [31] W.-B. Du, M.-Y. Zhang, Y. Zhang, X.-B. Cao, and J. Zhang, "Delay causality network in air transport systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 118, pp. 466–476, 2018.
- [32] Y. Xiao, Y. Zhao, G. Wu, and Y. Jing, "Study on delay propagation relations among airports based on transfer entropy," *IEEE Access*, vol. 8, pp. 97103–97113, 2020.
- [33] Y. Wang, H. Zheng, F. Wu, J. Chen, and M. Hansen, "A comparative study on flight delay networks of the USA and China," *Journal of Advanced Transportation*, vol. 2020, Article ID 1369591, 11 pages, 2020.
- [34] P. Mazzarisi, S. Zaoli, F. Lillo, L. Delgado, and G. Gurtner, "New centrality and causality metrics assessing air traffic network interactions," *Journal of Air Transport Management*, vol. 85, Article ID 101801, 2020.
- [35] B. Baspinar and E. Koyuncu, "A data-driven air transportation delay propagation model using epidemic process models," *International Journal of Aerospace Engineering*, vol. 2016, Article ID 4836260, 11 pages, 2016.
- [36] J. Mou, C. Liu, S. Chen, G. Huang, and X. Lu, "Temporal characteristics of the Chinese aviation network and their effects on the spread of infectious diseases," *Scientific Reports*, vol. 7, no. 1, Article ID 1275, 2017.
- [37] X. Wu, H. Yang, and S. Han, "Analysis of properties and delay propagation of air traffic based on complex network," *Hangkong Xuebao/Acta Aeronautica et Astronautica Sinica*, vol. 38, no. s1, Article ID 721473, 2017.
- [38] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 115, pp. 700–721, 1927.
- [39] N. Masuda, M. A. Porter, and R. Lambiotte, "Random walks and diffusion on networks," *Physics Reports*, vol. 716–717, pp. 1–58, 2017.
- [40] E. Estrada, "'Hubs-repelling' Laplacian and related diffusion on graphs/networks," *Linear Algebra and Its Applications*, vol. 596, pp. 256–280, 2020.
- [41] S. Li and X. Xu, "Vulnerability analysis for airport networks based on fuzzy soft sets: from the structural and functional perspective," *Chinese Journal of Aeronautics*, vol. 28, no. 3, pp. 780–788, 2015.
- [42] G. Ren, C. Lu, J. Zhu, and X. Liu, "Analyzing the topological characteristic and key nodes of Chinese air sector network," *International Journal of Modern Physics B*, vol. 33, no. 11, Article ID 1950100, 2019.
- [43] G. Ren, J. Zhu, and C. Lu, "A measure of identifying influential waypoints in air route networks," *PLoS One*, vol. 13, no. 9, Article ID e0203388, 2018.
- [44] G. Ren, J. Zhu, and C. Lu, "Identifying influential waypoints in air route networks based on network agglomeration relative entropy," *Chinese Journal of Physics*, vol. 57, pp. 382–392, 2019.
- [45] W. Du, B. Liang, G. Yan, O. Lordan, and X. Cao, "Identifying vital edges in Chinese air route network via memetic algorithm," *Chinese Journal of Aeronautics*, vol. 30, no. 1, pp. 330–336, 2017.
- [46] A. Voltes-Dorta, H. Rodríguez-Déniz, and P. Suañ-Sanchez, "Vulnerability of the European air transport network to major airport closures from the perspective of passenger delays: ranking the most critical airports," *Transportation Research Part A: Policy and Practice*, vol. 96, pp. 119–145, 2017.
- [47] L. Y. Lu, D. B. Chen, X. L. Ren, Q. M. Zhang, Y. C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Physics Reports-Review Section of Physics Letters*, vol. 650, pp. 1–63, 2016.
- [48] O. Lordan and R. Klophaus, "Measuring the vulnerability of global airline alliances to member exits," in *World Conference on Transport Research—WCTR 2016 (Transportation Research*

- Procedia*), F. Ulengin, K. Li, and M. Boltze, Eds., vol. 25, pp. 7–16, Elsevier Science Bv, Amsterdam, Netherlands, 2017.
- [49] W.-B. Du, X.-L. Zhou, O. Lordan, Z. Wang, C. Zhao, and Y.-B. Zhu, “Analysis of the Chinese airline network as multi-layer networks,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 89, pp. 108–116, 2016.
- [50] T. Verma, N. A. Araujo, and H. J. Herrmann, “Revealing the structure of the world airline network,” *Scientific Reports*, vol. 4, Article ID 5638, 2014.
- [51] M. Zanin and F. Lillo, “Modelling the air transport with complex networks: a short review,” *The European Physical Journal Special Topics*, vol. 215, no. 1, pp. 5–21, 2013.
- [52] M. E. J. Newman, “Measures and Metrics,” in *Networks: An Introduction*, pp. 167–240, Oxford University Press, Oxford, UK, 1st edition, 2010.
- [53] J. M. Anthonisse, “The rush in a directed graph,” in *Stichting Mathematisch Centrum. Mathematische Besliskunde* Stichting Mathematisch Centrum, Amsterdam, Netherlands, 1971.
- [54] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, “Comparing community structure identification,” *Journal of Statistical Mechanics-Theory and Experiment*, vol. 2005, no. 9, Article ID P09008, 2005.
- [55] S. N. Dorogovtsev and J. F. F. Mendes, “Evolution of networks,” *Advances in Physics*, vol. 51, no. 4, pp. 1079–1187, 2002.
- [56] 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.
- [57] L. Lu, T. Zhou, Q. M. Zhang, and H. E. Stanley, “The H-index of a network node and its relation to degree and coreness,” *Nature Communications*, vol. 7, Article ID 10168, 2016.
- [58] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, “The architecture of complex weighted networks,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.
- [59] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.
- [60] 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.
- [61] S. Li, X. Xu, F. Wang, and X. Wang, “Topological structure of US flight network based on complex network theory,” *Transactions of Nanjing University of Aeronautics and Astronautics*, vol. 32, no. 5, pp. 555–559, 2015.
- [62] O. Lordan and J. M. Sallan, “Analyzing the multilevel structure of the European airport network,” *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 554–560, 2017.
- [63] G. Bagler, “Analysis of the airport network of India as a complex weighted network,” *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 12, pp. 2972–2980, 2008.
- [64] X. Sun, V. Gollnick, and S. Wandelt, “Robustness analysis metrics for worldwide airport network: a comprehensive study,” *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 500–512, 2017.
- [65] R. Guimera, S. Mossa, A. Turttschi, and L. A. Amaral, “The worldwide air transportation network: anomalous centrality, community structure, and cities’ global roles,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 22, pp. 7794–7799, 2005.
- [66] G. Bianconi, P. Pin, and M. Marsili, “Assessing the relevance of node features for network structure,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 28, pp. 11433–11438, 2009.
- [67] T. Bian, J. T. Hu, and Y. Deng, “Identifying influential nodes in complex networks based on AHP,” *Physica A-Statistical Mechanics and Its Applications*, vol. 479, pp. 422–436, 2017.
- [68] A. Cardillo, M. Zanin, J. Gómez-Gardeñes, M. Romance, A. J. García del Amo, and S. Boccaletti, “Modeling the multi-layer nature of the European air transport network: resilience and passengers re-scheduling under random failures,” *The European Physical Journal Special Topics*, vol. 215, no. 1, pp. 23–33, 2013.
- [69] S. Hosseini, K. Barker, and J. E. Ramirez-Marquez, “A review of definitions and measures of system resilience,” *Reliability Engineering & System Safety*, vol. 145, pp. 47–61, 2016.
- [70] C. Nan and G. Sansavini, “A quantitative method for assessing resilience of interdependent infrastructures,” *Reliability Engineering & System Safety*, vol. 157, pp. 35–53, 2017.
- [71] R. Albert, H. Jeong, and A.-L. Barabási, “Error and attack tolerance of complex networks,” *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.
- [72] M. Janić, “Modelling the resilience, friability and costs of an air transport network affected by a large-scale disruptive event,” *Transportation Research Part A: Policy and Practice*, vol. 71, pp. 1–16, 2015.
- [73] S. Wandelt, X. Q. Sun, and X. B. Cao, “Computationally efficient attack design for robustness analysis of air transportation networks,” *Transportmetrica A-Transport Science*, vol. 11, no. 10, pp. 939–966, 2015.
- [74] S. Dunn and S. M. Wilkinson, “Increasing the resilience of air traffic networks using a network graph theory approach,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 90, pp. 39–50, 2016.
- [75] K.-C. Pien, K. Han, W. Shang, A. Majumdar, and W. Ochieng, “Robustness analysis of the European air traffic network,” *Transportmetrica A: Transport Science*, vol. 11, no. 9, pp. 772–792, 2015.
- [76] K.-C. Pien, *A linear programming approach for capacity estimation and robustness analysis of the european air traffic network*, Ph.D. dissertation, Department of Civil and Environmental Engineering, Imperial College London, London, UK, 2016.
- [77] S. Yoo and H. Yeo, “Evaluation of air transportation network resilience using adaptive capacity,” in *Proceedings of the 11th Traffic and Granular Flow Conference*, pp. 547–554, Delf, Netherlands, October 2015.
- [78] K. H. Thompson and H. T. Tran, “Operational perspectives into the resilience of the U.S. air transportation network against intelligent attacks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1503–1513, 2020.
- [79] Y. Wang, J. Zhan, X. Xu, L. Li, P. Chen, and M. Hansen, “Measuring the resilience of an airport network,” *Chinese Journal of Aeronautics*, vol. 32, no. 12, pp. 2694–2705, 2019.
- [80] L. Daqing, J. Yinan, K. Rui, and S. Havlin, “Spatial correlation analysis of cascading failures: congestions and blackouts,” *Scientific Reports*, vol. 4, Article ID 5381, 2014.
- [81] 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.
- [82] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, “Characterization of delay propagation in the US air-transportation network,” *Transportation Journal*, vol. 53, no. 3, pp. 330–344, 2014.

- [83] N. Pyrgiotis, K. M. Malone, and A. Odoni, “Modelling delay propagation within an airport network,” *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 60–75, 2013.
- [84] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, “Systemic delay propagation in the US airport network,” *Scientific Reports*, vol. 3, Article ID 1159, 2013.
- [85] J. J. Rebollo and H. Balakrishnan, “Characterization and prediction of air traffic delays,” *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 231–241, 2014.
- [86] X. Jiang, X. Wen, M. Wu, M. Song, and C. Tu, “A complex network analysis approach for identifying air traffic congestion based on independent component analysis,” *Physica A: Statistical Mechanics and Its Applications*, vol. 523, pp. 364–381, 2019.
- [87] C. Wang and X. Wang, “Airport congestion delays and airline networks,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 122, pp. 328–349, 2019.

Research Article

Assessing the Effectiveness of Mass Testing and Quarantine in the Spread of COVID-19 in Beijing and Xinjiang, 2020

Feng Li,^{1,2} Zhen Jin³ ,³ and Juan Zhang³

¹Data Science and Technology, North University of China, Taiyuan 030051, China

²College of Arts and Sciences, Shanxi Agricultural University, Taiyuan 030801, China

³Complex System Research Center, Shanxi University, Taiyuan, Shanxi 030006, China

Correspondence should be addressed to Zhen Jin; jinzhn@263.net

Received 24 January 2021; Revised 4 March 2021; Accepted 15 March 2021; Published 31 March 2021

Academic Editor: xiaoke xu

Copyright © 2021 Feng Li 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.

Coronavirus disease (COVID-19) cases and COVID-19-related deaths have been increasing worldwide since the outbreak in 2019. Before the mass vaccination campaign for COVID-19, the main methods for COVID-19 control in China were mass testing and quarantine. Based on the transmission mechanism of COVID-19, we constructed a dynamic model for COVID-19 transmission in two typical regions: Beijing and Xinjiang. We calculated the basic reproduction number R_0 , proved the global stability of COVID-19 transmission via the Lyapunov function technique, and introduced the final size. We assessed the effectiveness of mass testing and quarantine. Sensitivity analysis indicated that the more the people were tested per day, the larger is the quarantine proportionality coefficient, the earlier the source location was determined, and the better is the controlling effect. In addition, it was more effective to increase the coefficient of quarantine if the population density in the region was low. To eliminate the pandemic, the government has to expand testing and quarantine, requiring a large amount of continuous manpower, material, and financial resources. Therefore, new control measures should be developed.

1. Introduction

The sudden and unexpected outbreak of the coronavirus disease (COVID-19) pandemic is the most serious global public health emergency and has spread to more than 200 countries. COVID-19 is a contagious respiratory illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is a zoonotic virus. The symptoms of COVID-19 include fever or chills, cough, dyspnea, difficulty in breathing, fatigue, headache, nasal congestion or runny nose, muscle or body aches, sore throat, loss of smell or taste, nausea or vomiting, and diarrhea. Notably, some people become infected but do not develop any symptoms or feel unwell. Infectious droplets are released when someone with COVID-19 sneezes, coughs, or talks, which can settle in the mouth or nose of people who are nearby or possibly be inhaled into the lungs. COVID-19 is thought to spread mainly through close person-to-person contact (within approximately 6 feet). People who are infected but do not show symptoms may also spread the virus to others [1, 2].

In December 2019, COVID-19 was first reported in Wuhan, the capital of Hubei Province in central China. Early in the outbreak, many patients were reported to have a link with a large seafood and live animal market [3, 4]. On January 23, 2020, the Wuhan shutdown was implemented to limit the movement of people in and out of the city. Meanwhile, suspected and confirmed cases had been isolated, public transport had been suspended, schools and entertainment venues had been closed, public gatherings had been banned, and health checks and information had been widely disseminated [5, 6]. Through the lockdown, the spread of the virus was controlled. Residents returned to their normal lives 7 months after the city's strict lockdown was lifted.

In June and July 2020, COVID-19 affected Beijing and Xinjiang Uygur Autonomous Region (for Xinjiang), respectively, giving rise to the second wave of infections; however, this wave was localized within the regions and did not cause widespread transmission. The two regions curbed the rapid spread of COVID-19 cases by adopting strict

prevention and control measures. The details are as follows: on June 11, 2020, a new locally transmitted case was reported in Beijing. The Xinfadi wholesale market, which is the largest food wholesale market in Beijing, was identified as a possible source of infection within 22 hours. Most cases were linked to the market. The market was shut down, high-risk individuals were identified, and nearby communities were subjected to lockdown. The government ensured an increase in COVID-19 testing, including testing of suspected persons, and followed up on their status. With the implementation of targeted measures, Beijing controlled the surge in cases in less than a month. The number of daily infections returned to zero by July 7. The market was reopened to the public on August 16. On July 15, 2020, a 24-year-old woman from Urumqi (the capital of Xinjiang) tested positive for SARS-CoV-2, making her the first locally transmitted COVID-19 case in Xinjiang since February. Based on epidemiological investigations and laboratory testing, this round of outbreak was associated with a gathering, and most confirmed cases were reported in Urumqi. No newly confirmed or asymptomatic COVID-19 cases have been reported in Xinjiang since August 16. Before the mass vaccination campaign of COVID-19, there was successful management of the outbreak in Wuhan, which offered a valuable experience to those in Beijing and Xinjiang. Expanding the testing of the population and asking people from medium- and high-risk areas or who had close contact with confirmed or asymptomatic cases to quarantine themselves gated widespread community spread or spread across regions. These measures have effectively reduced the spread of the virus and prevented mass infections, thereby helping consolidate the efforts of COVID-19 control and prevention. These two outbreaks were quickly curbed. However, the time of pandemic control in Xinjiang was longer than that in Beijing.

With the massive global pandemic, it is natural that the outbreak emerges accidentally in other areas worldwide. There are geographical, economic, cultural, health care, and population density differences between people living in Beijing and Xinjiang. For instance, the population density in Beijing is about 1313 people per square kilometer, whereas it is 14.98 people per square kilometer in Xinjiang. Based on the initial spread of the infection, in Beijing, the source of the infection, that is, the market, was identified within 22 hours after the first case was confirmed. In Xinjiang, the epidemiological analysis showed that the virus source of this outbreak was from the same source of infection identified 13 days later. The other regions can learn from these two regions' experiences in containing the spread of the virus. Quantifying the effectiveness of these control measures is of crucial importance for cities preparing for rapid responses in the event of another outbreak; if asymptomatic cases are identified quickly, the source of the infection can be controlled and the infection rate can be reduced. Therefore, massive screening of asymptomatic cases and quarantine of susceptible people in communities are necessary.

The dynamic model is a useful method for identifying the key factors affecting disease outbreaks and evaluating the intervention strategies. Based on the transmission

mechanism, the model can dynamically predict future trends according to current information. The quantities used often include the basic or effective reproduction number and the final epidemic size [7–11]. A mathematical model is used to estimate the degree of the COVID-19 epidemic in cities worldwide [12–16]. The SEIR model is the most widely adopted model to study the COVID-19 outbreak in China and other countries [17–20]. Based on the SEIR model, some researchers have increased the number of states that have to follow the quarantine process [6, 21]. We analyzed the transmission of COVID-19 in two typical regions: Beijing and Xinjiang. Between the two regions, the outbreak occurred simultaneously and similar control measures were adopted, though there were significant differences, for instance, in the duration of confirming the source, the ability of mass testing, and proportional coefficient of quarantine.

In this study, we considered a dynamic model with intervention measures. The model is introduced in Section 2. In Section 3, the basic reproduction number, R_0 , a threshold quantity for the stability of equilibria, was calculated. Using the Lyapunov function, we proved that the disease-free equilibrium was globally asymptotically stable if $R_0 \leq 1$, whereas an endemic equilibrium was globally asymptotically stable if $R_0 > 1$. The final size was derived. In Section 4, we estimate the parameter values using Markov Chain Monte Carlo (MCMC) simulations. In Section 5, we explore the influence of mass testing and quarantine in Beijing and Xinjiang. The final size was evaluated for COVID-19 in Beijing and Xinjiang.

2. Model Formulation

The SIHRS_q model was used to estimate the COVID-19 epidemic in Beijing and Xinjiang. We divided the total population (N) into five groups: susceptible (S), unfound infected (I), hospitalized (H), removed (R , removed group includes recovered and death populations), and quarantine (S_q). Susceptible people were infected by asymptomatic infected people. Through a latent period, some infected people developed symptoms, went to the hospital, and were diagnosed, while others remained asymptomatic. Some asymptomatic people who were identified by nucleic acid testing were taken to the hospital. Therefore, hospitalized patients included symptomatic and asymptomatic individuals. Patients hospitalized were isolated and treated; thus, we considered that the susceptible people could not be infected by the hospitalized patients. Other asymptomatic people not diagnosed in the crowd recovered without treatment. Recovered patients included the treated and untreated patients. To minimize risks, susceptible people from medium- and high-risk areas or who had close contact with confirmed or asymptomatic cases were tested. If their results were negative on nucleic acid testing, they were prohibited from leaving the outbreak regions and quarantined. Therefore, the susceptible individuals were quarantined in proportion to the number of new cases hospitalized. Those quarantined remained till the outbreak was over in their regions. Hence, people quarantined did move toward susceptible people.

Based on the previous assumptions, the dynamic model was described by the following ordinary differential equations:

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\beta \frac{S}{N} I - \kappa_q \frac{S}{N} I + \mu N - \mu S, \\ \frac{dI(t)}{dt} = \beta \frac{S}{N} I - \alpha_1 I - \alpha_2 I - \mu I, \\ \frac{dH(t)}{dt} = \alpha_1 I - \delta H - \mu H, \\ \frac{dR(t)}{dt} = \alpha_2 I + \delta H - \mu R, \\ \frac{dS_q(t)}{dt} = \kappa_q \frac{S}{N} I - \mu S_q, \end{array} \right. \quad (1)$$

where $\alpha_2 = \gamma(1-\rho)(1-\lambda)$, $\alpha_1 = \rho(1-\lambda) + \lambda$, and $\kappa_q = \kappa(1-\beta)\alpha_1$, where β is the infection coefficient. λ denotes the discovery rate of infectious people. γ and δ are the recovery rates. $1/\mu$ is the rate of natural births and deaths, where ρ is the daily proportion of nucleic acid testing. With contact tracing, the uninfected individuals, who got in contact with new hospitalized cases, move to the quarantined compartment at a rate of κ_q . All parameters were nonnegative. A flow diagram of system (1) is shown in Figure 1. Additionally, the total population was included $N(t) = S(t) + I(t) + H(t) + R(t) + S_q(t)$. Adding all five equations in system (1), we obtain $(dN/dt) = 0$, for all $t \geq 0$, which implies $N(t) = N(0) > 0$ for all $t \geq 0$. Let the feasible region for system (1) be

$$\Omega = \left\{ (S, I, H, R, S_q) \in \mathbb{R}_+^5 : 0 < S + I + H + R + S_q \leq N \right\}. \quad (2)$$

And the initial condition is

$$\begin{aligned} S(0) &= S^0 \geq 0, & I(0) &= I^0 \geq 0, \\ H(0) &= H^0 \geq 0, & S_q(0) &= S_q^0 \geq 0, \\ & & R(0) &= R^0 \geq 0, \\ N(0) &= N^0 = S^0 + I^0 + H^0 + R^0 + S_q^0. \end{aligned} \quad (3)$$

Theorem 1. *The region Ω is positively invariant with respect to system (1).*

Proof. To show the positivity of each individual, we first show that $I(t) > 0$. We multiply both sides of the second equation of system (1) by $\Theta(t) = e^{-\int_0^t (\beta(S(\tau)/N)) d\tau + (\alpha_1 + \alpha_2 + \mu)t}$ and obtain

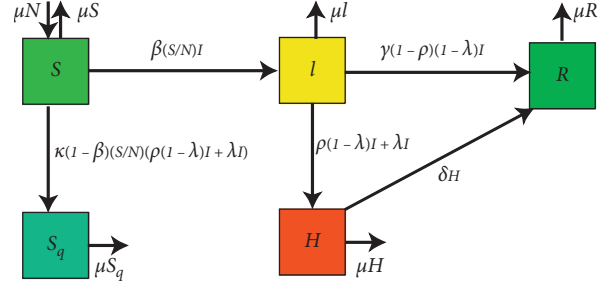


FIGURE 1: General transfer diagram of the model.

$$\Theta(t)I'(t) = \Theta(t) \left(\beta \frac{S(t)}{N} - (\alpha_1 + \alpha_2 + \mu) \right) I(t). \quad (4)$$

We obtain

$$I(t) = I^0 e^{\int_0^t (\beta(S(\tau)/N)) d\tau - (\alpha_1 + \alpha_2 + \mu)t}. \quad (5)$$

Since $I^0 \geq 0$, we obtain $I(t) \geq 0$ for all time $t > 0$. Now to show that $H(t) > 0$, we multiply both sides of the third equation of system (1) with $e^{(\delta+\mu)t}$ and obtain

$$e^{(\delta+\mu)t} H'(t) = e^{(\delta+\mu)t} (\alpha_1 I(t) - (\delta + \mu)H(t)). \quad (6)$$

We may write

$$\left(e^{(\delta+\mu)t} H(t) \right)' = e^{(\delta+\mu)t} \alpha_1 I(t), \quad (7)$$

on integrating from 0 to t and get

$$H(t) = e^{-(\delta+\mu)t} H^0 + \int_0^t e^{(\delta+\mu)(\tau-t)} \alpha_1 I(\tau) d\tau. \quad (8)$$

Since $H^0 \geq 0$ and $I(t) \geq 0$ for all times $t > 0$, we have $H(t) > 0$. We multiply both sides of the fourth equation of system (1) by $e^{\mu t}$. By integrating from 0 to t , we obtain

$$R(t) = e^{-\mu t} \left(R^0 + \int_0^t (\alpha_2 I(\tau) + \delta H(\tau)) d\tau \right). \quad (9)$$

Because $R^0 \geq 0$, $I(t) \geq 0$, and $H(t) \geq 0$ for all times $t > 0$, we have $R(t) > 0$. From the fifth equation of system (1) and $S = N - I - H - R - S_q$, we obtain

$$S_q'(t) = \kappa_q \frac{N - I(t) - H(t) - R(t)}{N} I(t) - \left(\kappa_q \frac{I(t)}{N} + \mu \right) S_q(t). \quad (10)$$

We multiply both sides of the above equation with $\Phi(t) = e^{\int_0^t (\kappa_q (I(\tau)/N)) d\tau + \mu t} > 0$ and get

$$\left(\Phi(t) S_q(t) \right)' = \kappa_q \Phi(t) \frac{N - I(t) - H(t) - R(t)}{N} I(t). \quad (11)$$

On integrating from 0 to t , we obtain that

$$S_q(t) = \Phi^{-1}(t) \left(S_q^0 + \int_0^t \left(\kappa_q \Phi(\tau) \frac{N - I(\tau) - H(\tau) - R(\tau)}{N} I(\tau) \right) d\tau \right). \quad (12)$$

Since $S_q^0 \geq 0$, $N - I(t) - H(t) - R(t) \geq 0$, and $I(t) \geq 0$ for all times $t > 0$, we have $S_q(t) > 0$. Similarly, to show that $S(t) > 0$, we multiply both sides of the first equation of system (1) by $\Psi(t) = e^{\int_0^t (\beta(I(\tau)/N) + \kappa_q(I(\tau)/N) + \mu) d\tau} > 0$ and obtain

$$\Psi(t)S'(t) = \Psi(t)\mu N - \Psi(t) \left(\beta \frac{I(t)}{N} + \kappa_q \frac{I(t)}{N} + \mu \right) S(t), \quad (13)$$

which can be rewritten as

$$(\Psi(t)S(t))' = \mu N \Psi(t). \quad (14)$$

On integrating from 0 to t , we obtain that

$$S(t) = \Psi^{-1}(t) \left(S^0 + \int_0^t \mu N \Psi(\tau) d\tau \right), \quad (15)$$

and thus $S(t) > 0$ for all times $t > 0$.

$N(t) = N$. Therefore, the region Ω is positively invariant. The solution enters Ω in an infinite time or $N(t)$ asymptotically approaches N . The region Ω attracts all solutions in R_+^5 [22]. \square

3. Dynamical Analysis

For the convenience of calculation, adding the equations in system (1), it follows that $N(t)$ is a first integral, and introducing the densities $s = (S/N)$, $i = (I/N)$, $h = (H/N)$, $s_q = (S_q/N)$, $r = (R/N)$, and $s + i + h + s_q + r = 1$, system (1) becomes

$$\begin{cases} \frac{ds(t)}{dt} = \mu - \beta si - \kappa_q si - \mu s, \\ \frac{di(t)}{dt} = \beta si - (\alpha_1 + \alpha_2)i - \mu i, \\ \frac{dh(t)}{dt} = \alpha_1 i - \delta h - \mu h, \\ \frac{dr(t)}{dt} = \alpha_2 i + \delta h - \mu r, \\ \frac{ds_q(t)}{dt} = \kappa_q si - \mu s_q. \end{cases} \quad (16)$$

3.1. Global Stability of the Disease-Free Equilibrium (DFE). By solving the equations of system (16) at a steady state, we obtain a disease-free equilibrium (DFE) $P^0 = (s^0, 0, 0, 0, 0)$,

where $s^0 = 1$. We can evaluate the basic reproduction number R_0 for system (16) following the next-generation matrix formulated by van den Driessche and Watmough [23]. For the epidemic model, R_0 shows that an average of each infected individual infects more than one individual. If $R_0 \leq 1$, the disease dies out. On the other hand, we may expect the disease to spread in the community if $R_0 > 1$. The basic reproduction number R_0 for system (16) is

$$R_0 = \frac{\beta}{\alpha_1 + \alpha_2 + \mu}. \quad (17)$$

Theorem 2. *The DFE of system (32) is locally asymptotically stable if $R_0 \leq 1$ and unstable if $R_0 > 1$.*

Proof. By linearizing system (16) around P^0 , the local stability of the DFE solution can be examined. Let $x = (i, h, r, s_q, s)$ and $f(x)$ denote the vector field of the system in (16). Jacobian matrix $J = (\partial f / \partial x)$ associated with this P^0 :

$$J|_{P^0} = \begin{pmatrix} \beta - \alpha_1 - \alpha_2 - \mu & 0 & 0 & 0 & 0 \\ \alpha_1 & -\delta - \mu & 0 & 0 & 0 \\ \alpha_2 & \delta & -\mu & 0 & 0 \\ \kappa_q & 0 & 0 & -\mu & 0 \\ -\beta - \kappa_q & 0 & 0 & 0 & -\mu \end{pmatrix}. \quad (18)$$

The characteristic equation for the above matrix is

$$(\lambda + \mu)^3 (\lambda + (\alpha_1 + \alpha_2 + \mu)(1 - R_0)) (\lambda + \delta + \mu) = 0. \quad (19)$$

We can see that all the eigenvalues have negative real parts if and only if $R_0 < 1$ which shows that the DFE P^0 is locally asymptotically stable. The DFE P^0 is unstable if $R_0 > 1$. \square

Theorem 3. *The DFE of system (16) is globally asymptotically stable in Ω whenever $R_0 < 1$.*

Proof. Consider the following Lyapunov function:

$$V(t) = i(t). \quad (20)$$

Then, if $R_0 < 1$,

$$\begin{aligned} V'(t) &= i'(t) \\ &= \beta si - (\alpha_1 + \alpha_2 + \mu)i \\ &\leq (\beta - \alpha_1 - \alpha_2 + \mu)i \\ &= (\alpha_1 + \alpha_2 - \mu)(R_0 - 1), \quad i \leq 0, \end{aligned} \quad (21)$$

where $V' = 0$ if and only if $i = 0$. The largest compact invariant set in $\{(s, i, h, r, s_q) \in \Omega: V' = 0\}$ is a singleton P^0 . The global stability of P^0 follows from LaSalle's invariance principle [24–26]. This establishes the theorem. \square

3.2. Global Stability of the Endemic Equilibrium. By solving the equations of system (16) in the steady state, there is an endemic equilibrium $P^* = (s^*, i^*, h^*, r^*, s_q^*)$ if $R_0 > 1$, where

$$\begin{aligned} s^* &= \frac{1}{R_0} > 0, \\ i^* &= \frac{(R_0 - 1)\mu}{\beta + \kappa_q} > 0, \\ h^* &= \frac{\alpha_1}{\delta + \mu} i^* > 0, \\ r^* &= \frac{\delta\alpha_1 + (\delta + \mu)\alpha_2}{(\delta + \mu)\mu} i^* > 0, \\ s_q^* &= \frac{(R_0 - 1)\kappa_q(\mu + \alpha_1 + \alpha_2)}{\beta(\beta + \kappa_q)} > 0. \end{aligned} \quad (22)$$

Now, we prove that P^* is globally asymptotically stable in the interior of Ω . We establish the following result.

Theorem 4. *The unique endemic equilibrium P^* is locally asymptotically stable if $R_0 > 1$.*

Proof. By linearizing system (16) about P^* , we obtain the Jacobian matrix:

$$J_{P^*} = \begin{pmatrix} \beta s^* - \alpha_1 - \alpha_2 - \mu & 0 & 0 & 0 & \beta i^* \\ \alpha_1 & -\delta - \mu & 0 & 0 & 0 \\ \alpha_2 & \delta & -\mu & 0 & 0 \\ \kappa_q s^* & 0 & 0 & -\mu & 0 \\ -(\beta + \kappa_q) s^* & 0 & 0 & 0 & -(\beta + \kappa_q) i^* - \mu \end{pmatrix}. \quad (23)$$

The characteristic equation for the above matrix is

$$(\lambda + M_1)(\lambda + M_2)(\lambda + \mu)(\lambda + \eta + \mu)(\lambda + \delta + \mu) = 0, \quad (24)$$

where

$$\begin{aligned} M_1 &= \frac{(\beta + \kappa_q)(\alpha_1 + \alpha_2 + \mu)i^*}{(\beta + \kappa_q)i^* + \mu}, \\ M_2 &= (\beta + \kappa_q)i^* + \mu. \end{aligned} \quad (25)$$

Hence, all the eigenvalues are negative real parts if $R_0 > 1$, which shows that P^* is locally asymptotically stable. \square

Theorem 5. *The endemic equilibrium P^* is globally asymptotically stable whenever $R_0 > 1$.*

Proof. We define the following Lyapunov function:

$$V = s^* \int_{s^*}^s \frac{\beta\tau - (\alpha_1 + \alpha_2 + \mu)}{\tau} d\tau + \int_{i^*}^i \frac{(\beta + \kappa_q)s^*\tau + \mu s^* - \mu}{\tau} d\tau, \quad (26)$$

where $\alpha_1 + \alpha_2 + \mu = \beta s^*$. By calculating the time derivative of V along the solution of system (16), we obtain

$$V' = -\frac{\mu\beta(s - s^*)^2}{s} \leq 0, \quad (27)$$

for all $(s, i) \in \Omega$. Therefore, (s^*, i^*) is globally asymptotically stable.

Using the previous conclusion,

$$h(t) = e^{-(\delta + \mu)t} h^0 + \int_0^t e^{(\delta + \mu)(\tau - t)} \alpha_1 i(\tau) d\tau. \quad (28)$$

We derive

$$\lim_{t \rightarrow \infty} h(t) = \lim_{t \rightarrow \infty} \frac{\alpha_1 i(t)}{\delta + \mu} = i^*. \quad (29)$$

Similarly, we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} r(t) &= r^*, \\ \lim_{t \rightarrow \infty} s_q(t) &= s_q^*. \end{aligned} \quad (30)$$

According to LaSalle's invariance principle [27, 28], the endemic equilibrium P^* is unique and globally asymptotically stable in the interior of Ω if $R_0 > 1$. The theorem is proven. \square

3.3. Final Size. The final size of the epidemic is the number of individuals ultimately becoming infected, which is the total number of cases during the entire outbreak [29, 30]. In analyzing system (16), we adopt the conventions that, for an arbitrary continuous function $x(t)$ with nonnegative components:

$$x^\infty = \lim_{t \rightarrow \infty} x. \quad (31)$$

Due to the small-time span, we considered natural birth and death to be zero.

Letting $\mu = 0$ in system (16), it becomes

$$\begin{cases} \frac{ds(t)}{dt} = -\beta si - \kappa_q si, \\ \frac{di(t)}{dt} = \beta si - (\alpha_1 + \alpha_2)i, \\ \frac{dh(t)}{dt} = \alpha_1 i - \delta h, \\ \frac{dr(t)}{dt} = \alpha_2 i + \delta h, \\ \frac{ds_q(t)}{dt} = \kappa_q si. \end{cases} \quad (32)$$

The disease basic reproductive number is

$$R_0 = \frac{\beta}{\alpha_1 + \alpha_2}. \quad (33)$$

We observe that

$$\begin{aligned} (s(t) + s_q(t))' &< 0, \\ (s(t) + s_q(t) + i(t))' &< 0, \\ (s(t) + s_q(t) + i(t) + h(t))' &< 0, \\ r'(t) &> 0, \end{aligned} \quad (34)$$

for all $t > 0$, while we know that $s(t), s_q(t), i(t), h(t) > 0$ and $0 \leq r(t) < N$; hence, s^∞ and r^∞ exist, and we have $i^\infty = 0$, $h^\infty = 0$, and $s_q^\infty = 0$. Adding $s, s_q,$ and i equations to system (32), we obtain

$$(s + s_q + i)' = -(\alpha_1 + \alpha_2)i. \quad (35)$$

Integration of the equation with respect to t from 0 to ∞ gives

$$\int_0^\infty idt = \frac{1}{\alpha_1 + \alpha_2} (s^0 - s^\infty + s_q^0 + i^0). \quad (36)$$

Adding sand s_q equations to system (32),

$$(s + s_q)' = -\beta si. \quad (37)$$

From $(di/d(s + s_q))$, we obtain that

$$\begin{aligned} di &= -d(s + s_q) + \frac{1}{R_0} \left(\frac{ds}{s} + \frac{ds_q}{s} \right) \\ &= -d(s + s_q) + \frac{1}{R_0} \frac{ds}{s} + \frac{\kappa_q}{R_0} i. \end{aligned} \quad (38)$$

An epidemic ends when no infection is found; that is, $i^\infty = 0$, $h^\infty = 0$. The integration of the above equation yields

$$\ln\left(\frac{s^\infty}{s^0}\right) = -R_0 \left(1 + \frac{\kappa_q}{\beta}\right) (s^0 - s^\infty + s_q^0 + i^0). \quad (39)$$

We now derive the equation for the final epidemic size, defined by $Y = s^0 - s^\infty$. This yields an implicit relation:

$$Y = s^0 e^{-R_0 (1 + (\kappa_q/\beta)) (Y + s_q^0 + i^0)}. \quad (40)$$

4. Estimation of Epidemiological Parameters

We obtained data on COVID-19 cases in Beijing and Xinjiang from the National Health Commission of The People's Republic of China. The data included the cumulative and new infections of confirmed, suspected, death, cured, and asymptomatic cases; the cumulative number of close contacts; and hospitalization time [31].

In China, the per-day rate of natural birth and death is $\mu = 0.01094/365$. Suppose that the recovery rate of asymptomatic cases is $\gamma = 1/14$. Beijing is a metropolitan area with a population of 21.536 million ($N = 2.1536 \times 10^7$). From June 11 to July, 9,335 confirmed local cases and 49

asymptomatic cases had been reported, and no deaths were reported of the cases in Beijing. According to the data in Beijing, the mean time from infection to being tested positive for the virus was 10.8 days $\lambda = (1/10.8)$, and the treatment time was 10 days $\delta = (1/10)$. The testing capacity was approximately 0.3 million people per day $\rho = 3 \times 10^5/N$. The initial values on June 11 in Beijing were $S(0) = 2.1536 \times 10^7$, $R(0) = 0$ and $H(0) = 1$.

The population in Xinjiang is 24.87 million ($N = 2.487 \times 10^7$). Since the outbreak of the virus on July 15, 828 confirmed COVID-19 cases and 391 asymptomatic cases have been reported in Xinjiang, mainly in Urumqi. No newly confirmed COVID-19 case had been reported in Xinjiang since August 16. The treatment time was 11 days $\delta = 1/11$. From July 15 to 19, about 0.24 million people who had close contact with infected people and asymptomatic carriers had got the nucleic acid test, where the testing capacity was 0.06 million people per day ($\rho_1 = 6 \times 10^4/N$). The mean time from infection to being tested positive for the virus was 14 days ($\lambda_1 = 1/14$). On July 18, a citywide, free nucleic acid test was launched in Xinjiang. The testing capacity was up to 0.5 million people per day ($\rho_2 = 5 \times 10^5/N$) since July 20, the meantime from getting infected to being tested positive for the virus was shortened ($\lambda_2 = 1/7$). The initial values on June 11 in Xinjiang were $S(0) = 2.487 \times 10^7$, $R(0) = 0$, and $H(0) = 1$.

Based on the mathematical model, the cumulative numbers of confirmed cases and identified asymptomatic cases were determined in Beijing and Xinjiang. We used the MCMC method to fit the model and adopted an adaptive Metropolis-Hastings algorithm to carry out the MCMC procedure. The algorithm was run for 30,000 iterations with a burn-in of the first 10,000 iterations, and the Geweke convergence diagnostic method was employed to assess the convergence of chains [32, 33]. We could acquire the parameter values for β and κ , and the initial values for S_q^0 and I^0 in the two regions. The mean values and 95% confidence intervals (95% CI) are presented in Table 1.

The infection coefficient in Beijing ($\beta = 0.0640$) is lower than that in Xinjiang ($\beta = 0.1522$). S_q^0 in Beijing is greater than that in Xinjiang because the population density in Beijing (approximately 1313 people per square kilometer) is higher than that in Xinjiang (approximately 14.98 people square kilometer). I^0 in Beijing is lower than that in Xinjiang because the source of infection in Beijing was identified quicker than that in Xinjiang. The time evolution of the infected cases and comparison with the cumulative number of confirmed cases and identified asymptomatic cases are shown in Figure 2. Evidently, the theoretical prediction is nearly in agreement with the actual data, which also validates the accuracy of the model. The final sizes of the pandemic in Beijing and Xinjiang were $S(\infty) = 2.1535137 \times 10^7$, $R(\infty) = 862$ and $S(\infty) = 2.4866980 \times 10^7$, $R(\infty) = 3019$, respectively. The final scale in Beijing is smaller.

5. Spread of COVID-19 and the Effectiveness of Interventions in Beijing and Xinjiang

The basic reproduction numbers in Beijing and Xinjiang were 0.3783 and 1.0882, respectively. The basic reproductive

TABLE 1: Parameter estimation for COVID-19 in Beijing and Xinjiang.

| Parameter | Beijing | | Xinjiang | | Source |
|---------------|----------|----------------------|----------|----------------------|--------|
| | Mean | 95% CI | Mean | 95% CI | |
| β | 0.0640 | [0.0640, 0.0641] | 0.1522 | [0.1520, 0.1524] | MCMC |
| κ | 99.2342 | [93.4580, 105.8892] | 83.3883 | [83.3727, 83.4038] | MCMC |
| ρ | 80.0866 | [80.0864, 80.0868] | 1.8793 | [1.8790, 1.8796] | MCMC |
| τ_{as}^d | 344.4526 | [344.1625, 344.7428] | 449.7552 | [447.4628, 451.0475] | MCMC |

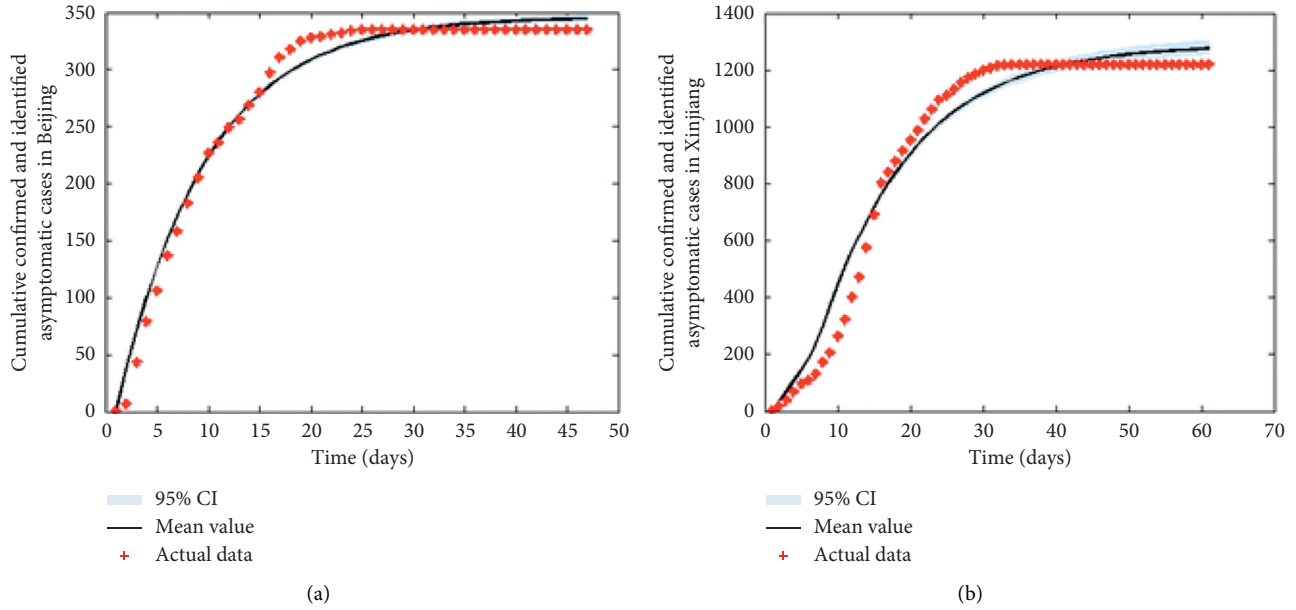


FIGURE 2: Fitting results of the estimated cumulative number of confirmed and identified asymptomatic COVID-19 cases with its actual reported numbers in Beijing and Xinjiang.

number in Beijing was less than one, and it was verified that the disease dies out. The basic reproductive number in Xinjiang was greater than one; however, with the increasing expansion of the nucleic acid testing, the outbreak was curbed. To draft policies for control measures, we analyzed the cumulative number of confirmed and identified asymptomatic COVID-19 cases for different values of κ (Figure 3). The analysis showed that, to control the spread of COVID-19 (the smaller the final size, the shorter the duration), a larger coefficient of quarantine is necessary. In addition, as the population density in Xinjiang is less than that in Beijing, it is effective to increase the coefficient of quarantine in Xinjiang. We provided a sensitivity analysis of the basic production number R_0 with the parameters β and ρ in Beijing and Xinjiang (Figure 4). Apparently, improving the testing capacity ρ can decrease the basic reproduction number R_0 . When the testing capacity reaches a certain level, it will make the basic reproduction number R_0 less than one (Figure 4). In Xinjiang, the basic reproduction number $R_0 > 1$ was equal to $\rho \geq 415600/N$ if the other parameters are fixed. This implies that the infection could not persist over time in Xinjiang if the number of people tested was greater than 415,600 per day. Without mass vaccination campaigns, mass testing may be an effective prevention measure. Early

detection of new cases is a useful control strategy for COVID-19.

6. Conclusion and Discussion

In this study, we constructed a mathematical model for COVID-19 transmission in Beijing and Xinjiang in 2020. There are differences in geographical, economic, cultural, health care, and lifestyles of people living in the two regions. The population in Xinjiang is similar to that in Beijing; however, the population density in Xinjiang is 100 times less than that in Beijing. The outbreak occurred in Beijing and then in Xinjiang. Similar measures (mass testing and quarantine) were adopted in the two regions, though the time to confirm the source, the ability of mass testing, and the proportional coefficient of quarantine were different. To explore effective control and prevention measures for COVID-19, we proposed a deterministic model with mass testing and quarantine to describe the spread of COVID-19 in China. First, the basic reproduction number R_0 of the model was provided. Then, it was found that the model had two nonnegative equilibria: the DFE and endemic equilibrium. Through the analysis of the model, it was found that the global behavior of the system was completely determined

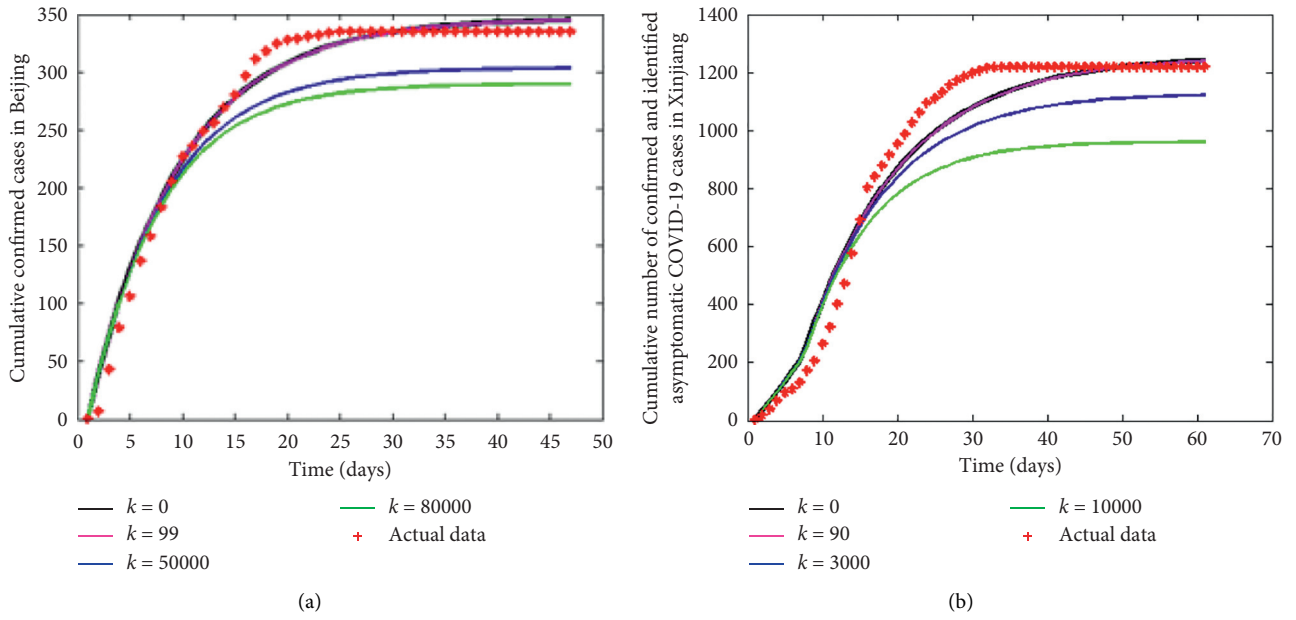


FIGURE 3: Cumulative number of confirmed and identified asymptomatic COVID-19 cases on κ in (a) Beijing and (b) Xinjiang.

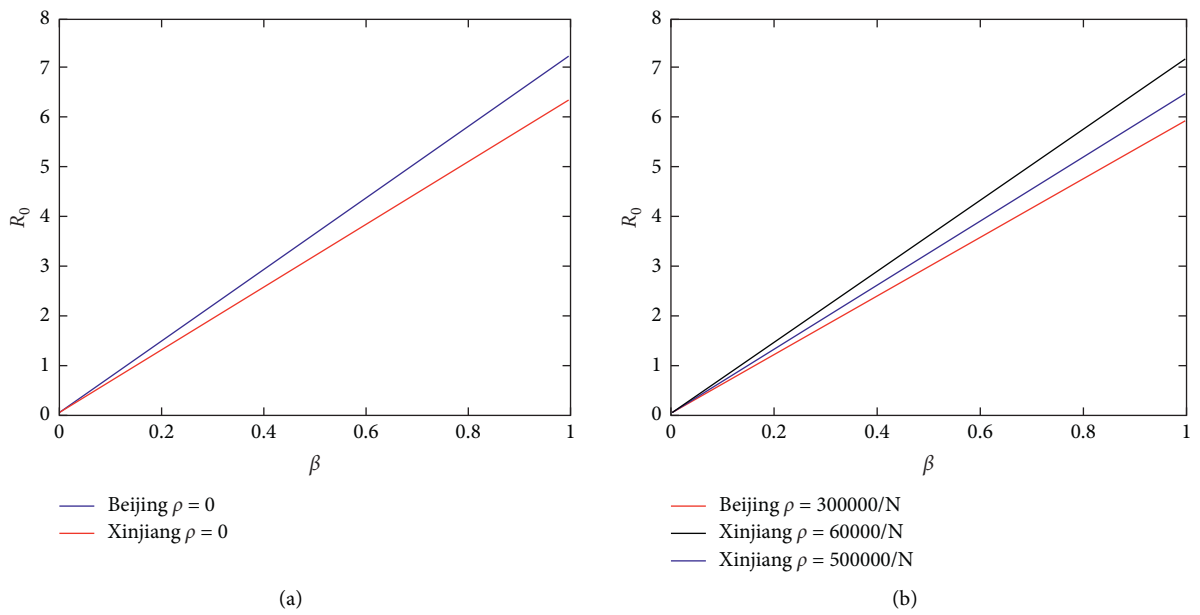


FIGURE 4: Continued.

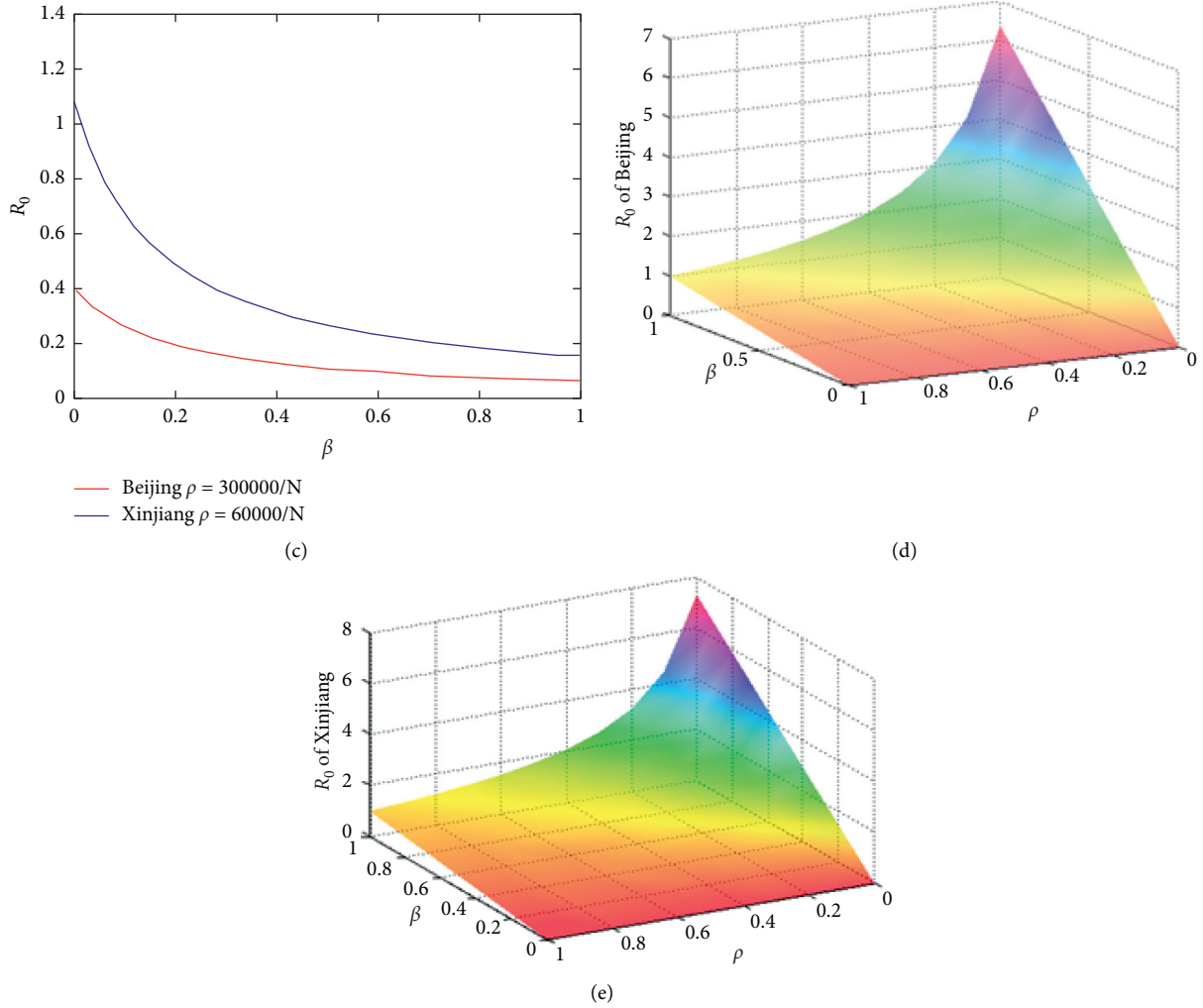


FIGURE 4: (a) R_0 in terms of β . (b) R_0 in terms of ρ . (c) R_0 in terms of ρ . (d) R_0 in terms of β and ρ for Beijing. (e) R_0 in terms of β and ρ for Xinjiang.

by the size of the basic reproduction number R_0 , that is, the DFE was globally asymptotically stable if $R_0 < 1$, while an endemic equilibrium existed uniquely and was globally asymptotically stable if $R_0 > 1$. Based on mathematical analysis, data fitting, and sensitivity analysis, we revealed the effects of mass testing and quarantine. Without the mass vaccination campaign of COVID-19, it was found that the higher the mass testing is, the earlier the source of infection was determined for effectively controlling the spread. In addition, a larger coefficient of quarantine was necessary to control the spread of COVID-19. It was more effective in increasing the coefficient of quarantine if the population density of the region was less. If the government implements mass testing and quarantine, it must devote a large amount of continuous manpower, material, and financial resources to eliminate the epidemic. With the development of the vaccine for COVID-19, more people are vaccinated in a community, which makes it possible for eliminating the spread of COVID-19.

In the future, we plan to study the mass vaccination campaign for COVID-19 in the whole of China.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request (email: lifeng@sxau.edu.cn).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61873154), Shanxi Key Laboratory

(201705D111006), and Shanxi Scientific and Technology Innovation Team (201705D1511172) Grant.

References

- [1] S. Ruan, "Likelihood of survival of coronavirus disease 2019," *The Lancet Infectious Diseases*, vol. 20, no. 6, pp. 630–631, 2020.
- [2] "Centers for Disease Control and Prevention," <http://www.nhc.gov.cn/>.
- [3] N. Zhu, D. Zhang, W. Wang et al., "A novel coronavirus from patients with pneumonia in China, 2019," *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020.
- [4] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020.
- [5] L. Pang, S. Liu, X. Zhang, T. Tian, and Z. Zhao, "Transmission dynamics and control strategies of COVID-19 in Wuhan, China," *Journal of Biological Systems*, vol. 415, pp. 1–18, 2020.
- [6] B. Tang, X. Wang, Q. Li et al., "Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions," *Journal of Clinical Medicine*, vol. 9, no. 2, pp. 462–475, 2020.
- [7] H. L. Smith, L. Wang, and M. Y. Li, "Global dynamics of an SEIR epidemic model with vertical transmission," *Siam Journal on Applied Mathematics*, vol. 62, pp. 58–69, 2001.
- [8] H. W. Hethcote, "The mathematics of infectious diseases," *Siam Review*, vol. 42, no. 4, pp. 599–653, 2000.
- [9] J. M. Heffernan, R. J. Smith, and L. M. Wahl, "Perspectives on the basic reproductive ratio," *Journal of the Royal Society Interface*, vol. 2, no. 4, pp. 281–293, 2005.
- [10] A. Korobeinikov, "Global properties of basic virus dynamics models," *Bulletin of Mathematical Biology*, vol. 66, no. 4, pp. 879–883, 2004.
- [11] P. Magal, O. Seydi, and G. Webb, "Final size of an epidemic for a two-group SIR model," *Siam Journal on Applied Mathematics*, vol. 76, no. 5, pp. 2042–2059, 2016.
- [12] R. Subramanian, Q. He, and M. Pascual, "Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology and testing capacity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, Article ID e2019716118, 2021.
- [13] Q. Li, N. L. Bragazzi, S. Tang, Y. Xiao, B. Tang, and X. Wang, "Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions," *Journal of Clinical Medicine*, vol. 9, pp. 462–475, 2020.
- [14] C. Yang and J. Wang, "Modeling the transmission of COVID-19 in the US - a case study," *Infectious Disease Modelling*, vol. 6, pp. 195–211, 2021.
- [15] J. Arino and S. Portet, "A simple model for COVID-19," *Infectious Disease Modelling*, vol. 5, pp. 309–315, 2020.
- [16] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," *Chaos Solitons Fractals*, vol. 134, Article ID 109761, 2020.
- [17] T. Chen, J. Rui, Q. Wang, Z. Zhao, J. Cui, and L. Yin, "A mathematical model for simulating the phase-based transmissibility of a novel coronavirus," *Infectious Diseases of Poverty*, vol. 9, pp. 24–32, 2020.
- [18] Y. Chen, J. Cheng, Y. Jiang, and K. Liu, "A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification," *Journal of Inverse and Ill-Posed Problems*, vol. 28, no. 2, pp. 243–250, 2020.
- [19] G.-Q. Sun, S.-F. Wang, M.-T. Li et al., "Transmission dynamics of COVID-19 in Wuhan, China: effects of lockdown and medical resources," *Nonlinear Dynamics*, vol. 101, no. 3, pp. 1981–1993, 2020.
- [20] L. Mangoni and M. Pistilli, *Epidemic Analysis of COVID-19 in Italy by Dynamical Modeling*, Social Science Electronic Publishing, Rochester, NY, USA, 2020.
- [21] B. Tang, N. L. Bragazzi, Q. Li, S. Tang, Y. Xiao, and J. Wu, "An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV)," *Infectious Disease Modelling*, vol. 5, pp. 248–255, 2020.
- [22] M. O. Ibrahim, "A mathematical model for the epidemiology of tuberculosis with estimate of the basic reproduction number," *IOSR Journal of Mathematics*, vol. 5, pp. 46–52, 2013.
- [23] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Mathematical Biosciences*, vol. 180, no. 1–2, pp. 29–48, 2002.
- [24] N. P. Bhatia and G. P. Szeg'o, *Dynamical Systems: Stability Theory and Applications*, Vol. 35, Springer, Berlin, Germany, 1967.
- [25] J. P. LaSalle, "Stability of nonautonomous systems," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 1, no. 1, pp. 83–90, 1976.
- [26] M.-T. Li, G.-Q. Sun, Y.-F. Wu, J. Zhang, and Z. Jin, "Transmission dynamics of a multi-group brucellosis model with mixed cross infection in public farm," *Applied Mathematics and Computation*, vol. 237, pp. 582–594, 2014.
- [27] J. P. LaSalle, *The Stability of Dynamical Systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1976.
- [28] L. Salle, Joseph, S. Lefschetz, and R. C. Alverson, *Stability by Liapunov's Direct Method with Applications*, Academic Press, Cambridge, MA, USA, 1961.
- [29] J. Ma, D. J. D. Earn, and J. D. David, "Generality of the final size formula for an epidemic of a newly invading infectious disease," *Bulletin of Mathematical Biology*, vol. 68, no. 3, pp. 679–702, 2006.
- [30] L. Davidson, "An introduction to mathematical biology," *Development*, vol. 138, pp. 5269–5270, 2011.
- [31] National Health Commission of the People's Republic of China, *Daily Report about 2019-nCoV*, National Health Commission, Beijing, China, 2020, <https://www.cdc.gov/coronavirus/2019-nCoV/index.html>.
- [32] H. Haario, M. Laine, A. Mira, and E. Saksman, "DRAM: efficient adaptive MCMC," *Statistics and Computing*, vol. 16, no. 4, pp. 339–354, 2006.
- [33] D. Sinha, "Markov chain Monte Carlo: Stochastic simulation for Bayesian inference," *Journal of the American Statistical Association*, vol. 104pp. 422–423, 2nd edition, 2009.

Research Article

Coupled Dynamic Model of Resource Diffusion and Epidemic Spreading in Time-Varying Multiplex Networks

Ping Huang,^{1,2,3} Xiao-Long Chen,⁴ Ming Tang,^{5,6} and Shi-Min Cai ^{1,2,3}

¹School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

²Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China

³Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 610054, China

⁴School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China

⁵School of Mathematical Sciences, Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China

⁶Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China

Correspondence should be addressed to Shi-Min Cai; shimin.cai81@gmail.com

Received 27 October 2020; Revised 3 March 2021; Accepted 19 March 2021; Published 28 March 2021

Academic Editor: Ye Wu

Copyright © 2021 Ping Huang 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.

In the real world, individual resources are crucial for patients when epidemics outbreak. Thus, the coupled dynamics of resource diffusion and epidemic spreading have been widely investigated when the recovery of diseases significantly depends on the resources from neighbors in static social networks. However, the social relationships of individuals are time-varying, which affects such coupled dynamics. For that, we propose a coupled resource-epidemic (RNR-SIS) dynamic model (coupled model for short) on a time-varying multiplex network to synchronously simulate the resource diffusion and epidemic spreading in dynamic social networks. The equilibrium analysis of the coupled model is conducted in a general scenario where the resource generation varies between susceptible and infected states and the recovery rate changes between resourceful and noresource states. By using the microscopic Markov chain approach and Monte Carlo simulations, we determine a probabilistic framework of the intralayer and interlayer dynamic processes of the coupled model and obtain the outbreak threshold of epidemic spreading. Meanwhile, the experimental results show the trivially asymmetric interactions between resource diffusion and epidemic spreading. They also indicate that the stronger activity heterogeneity and the larger contact capacity of individuals in the resource layer can more greatly promote resource diffusion, effectively suppressing epidemic spreading. However, these two individual characters in the epidemic layer can cause more resource depletion, which greatly promotes epidemic spreading. Furthermore, we also find that the contact capacity finitely impacts the coupled dynamics of resource diffusion and epidemic spreading.

1. Introduction

Most real networks are not isolated and the spreading dynamic processes on such real networks described by classical models such as the susceptible-infected-susceptible model (SIS) [1, 2] and susceptible-infected-recovery model (SIR) [3] may be interrelated and interactive with each other [4–6]. In epidemic spreading, a variety of factors can greatly influence the spread of diseases, such as the particular individuals' birth rate [7], immunity [8, 9], public policy [10, 11], vaccination [12], awareness [13–19], and resource [20, 21].

For example, Brandeau [20] et al. determined the optimal allocation of a limited resource for epidemic control among independent populations, and Chen et al. [21] found that any additional quantity of resources above the critical level would only bring marginal benefits to the containment. These works focus more on how public resource investment affects the spread of disease. However, in a general scenario, many patients tend to depend on the individual resources from their direct or indirect neighbors in social networks because of the limitation of public resource investment. Thus, some researchers naturally investigate the interactions between individual resources and epidemic dynamics. They

assume that resource diffusion and epidemic spreading occur on a two-layer interdependent network because resources usually propagate through social links and diseases spread through physical contacts. They find that individual resources play a crucial role in controlling or mitigating epidemic spreading by using an appropriate strategy of resource allocation [22–27].

Although many works based on contact network models have been done on the resource allocation for epidemic control, most of them assume that contacts are static through time. However, contact patterns are not static [28–31] and abundant close contacts in heterogeneous social networks promote both epidemic outbreak and spreading [32–35]. In recent years, the activity-driven (AD) temporal network model has attracted a large amount of attention for considering the dynamic connectivity patterns in real networks [19, 36–39]. In such network models, each node is assigned with an activity potential that determines its probability to create pairwise interactions with others randomly selected at each time.

Furthermore, some theoretical methods describe the evolution of the dynamic spreading process and analyze the corresponding stylized characters. Two of the widely used approaches are heterogeneous mean-field (HMF) [40–42] and the microscopic Markov chain approach (MMCA) [43]. MMCA can deal with individual-level infection dynamics much better. This approach is first used to derive the expression of the outbreak threshold of epidemic spreading on epidemic-awareness multiplex networks [44, 45]. Very recently, the MMCA is also expanded to understand the interaction between awareness propagation and epidemic spreading on static multiplex networks [46, 47] and time-varying multiplexes [19, 48].

In real life, healthy individuals can earn resources (e.g., money) or strive for personal resources. However, patients quickly lose the ability to generate resources. For solving the problem of deficiency of individual resources, patients commonly seek resources from their friends or relatives and even use the crowdfunding platform (e.g., GoFundMe and Waterdrop) to obtain resources provided by healthy individuals. As the resource diffusion and epidemic spreading on dynamic social networks synchronously happen in real society, the present work uses the AD model to construct a generic time-varying multiplex network representing the active contacts (or social relationships) of social networks. Because the dissemination of resources relies on the spreading of news on the crowdfunding platform, we use the Resourceful-Noresource-Resourceful (RNR) model (i.e., a SIS-like model) to describe resource diffusion. The evolutionary process of epidemic spreading is described by the SIS model. These two dynamic models are coupled into the RNR-SIS model by interlayer links of the time-varying multiplex network better to understand the interactions between resource diffusion and disease spreading. Meanwhile, considering the similarity between awareness propagation and resource diffusion, we make efforts to apply the MMCA for constructing a theoretical framework of the coupled resource-epidemic dynamic model (i.e., coupled RNR-SIS model) on a time-varying multiplex network.

Theoretically, the equilibrium analysis based on MMCA determines a probabilistic framework of the coupled model and theoretical outbreak thresholds involving different parameters. By performing extensive numerical simulations, we obtain the outbreak thresholds of epidemic spreading via variability measurement, which well agree with the theoretical analysis. Moreover, the experimental results also show that stronger activity heterogeneity and larger contact capacity of individuals greatly influence the coupled dynamics of resource diffusion and epidemic spreading. However, the influence of small contact capacity is finite.

Our paper is organized as follows: in Section 2, we describe the coupled RNR-SIS model on a time-varying multiplex network. In Section 3, we use MMCA to propose a probabilistic description of the dynamical processes and derive the theoretical expression of the outbreak threshold. In Section 4, we compare numerical simulations with theoretical analysis and investigate the influences of different parameters on resource diffusion and epidemic spreading. In Section 5, we conclude our work.

2. Coupled Resource-Epidemic Model on a Time-Varying Multiplex Network

The coupled RNR-SIS model on a time-varying multiplex network is depicted in Figure 1. In this model, individuals are mapped into the nodes of an AD temporal two-layer network. The individual resources diffuse on the upper layer (i.e., resource layer), and the spread of disease occurs on the lower layer (i.e., epidemic layer). In the dynamic process of resource diffusion (or epidemic spreading), the nodes are alternatively changing from the noresource (or susceptible) to resourceful (or infected) states according to resource rate diffusion λ (or epidemic transmission rate β) and conversely with resource loss rate δ (or epidemic recovery rate μ). λ is parallel with β . For example, the possibility is that you will give a hand when you browse information about fundraising. Thus, in the coupled dynamic process, the nodes are divided into four categories, noresource and susceptible (NS), resourceful and susceptible (RS), noresource and infected (NI), and resourceful and infected (RI). We assume the trivially asymmetric interactions between resource diffusion and epidemic spreading because the individual resources effectively affect the recovery of disease, but the spread of disease only affects the generation of individual resources. Thus, we further refine the epidemic recovery rate into μ^N and μ^R according to the node state in the epidemic layer. Obviously, the infected nodes with the resource can recover faster than those without resources, which indicates $\mu^N = \xi\mu^R$ ($0 \leq \xi \ll 1$). Table 1 shows the detailed definition of parameters in the coupled model. Most of the parameters are probabilities and rates, and their domain is limited in the interval $[0, 1]$.

More concretely, we construct the AD networks of N nodes for each layer regarding different parameters. Each node i is assigned with an activity level (i.e., firing rate) $a_i = \eta_1 x_i$ in the resource layer and $b_i = \eta_2 y_i$ in the epidemic layer, where η_i is a rescaling factor to make the average number of active nodes per unit time be $\eta_1 \langle x \rangle N$ and

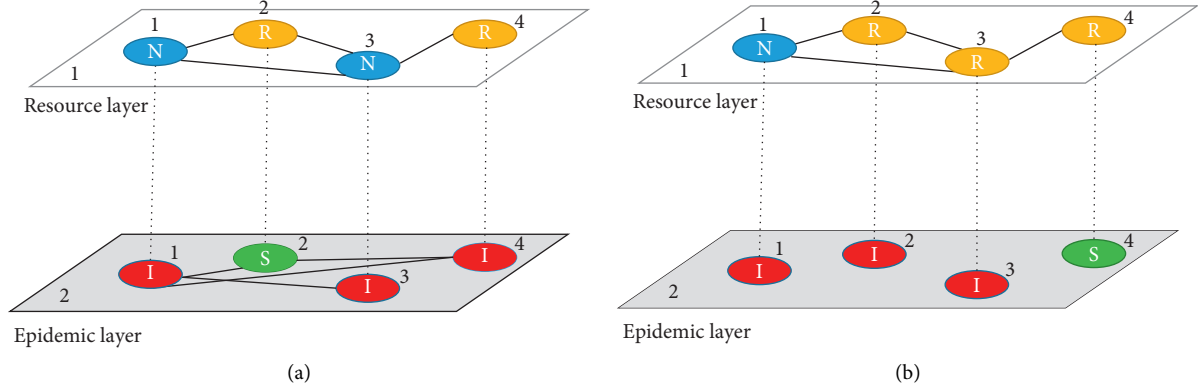


FIGURE 1: The sketch of coupled RNR-SIS model on an AD temporal two-layer network. The nodes indicate individuals in real society, and the edges depict dynamic contacts and social relationships of individuals in social networks. Individual resources diffuse on the upper layer (i.e., resource layer), leading to the node in a resourceful or noresource state. The spread of disease occurs on the lower layer (i.e., epidemic layer), leading to the node in a susceptible or infected state. The coupled dynamic process evolves from (a) times t_1 to (b) $t_1 + \Delta t$, where node 2 is infected from its neighbors, node 4 recovers in the epidemic layer, and node 3 changes from the noresource to resourceful state in the resource layer.

TABLE 1: Notions of parameters.

| Symbol | Domain | Definition |
|-------------|-----------------|---|
| N | $[0, \infty]$ | Number of nodes in network |
| t | $[0, \infty]$ | Time step |
| δ | $[0, 1]$ | Resource loss rate |
| λ | $[0, 1]$ | Resource diffusion rate |
| β | $[0, 1]$ | Epidemic transmission rate |
| μ | $[0, 1]$ | Epidemic recovery rate |
| μ^N | $[0, 1]$ | Epidemic recovery rate of the nodes in noresource state |
| μ^R | $[0, 1]$ | Epidemic recovery rate of the nodes in resourceful state |
| m_1 | $[0, \infty]$ | Contact capacity of active nodes in the resource layer |
| m_2 | $[0, \infty]$ | Contact capacity of active nodes in the epidemic layer |
| a_i | $[0, 1]$ | Activity level of node i in the resource layer |
| b_i | $[0, 1]$ | Activity level of node i in the epidemic layer |
| η_1 | $[0, \infty]$ | Rescaling factor of resource layer |
| η_2 | $[0, \infty]$ | Rescaling factor of epidemic layer |
| x_i | $[\epsilon, 1]$ | Activity potential of node i in the resource layer |
| y_i | $[\epsilon, 1]$ | Activity potential of node i on epidemic layer |
| ϵ | $(0, 1]$ | The lower cutoff of activity potential |
| $F_1(x)$ | — | Distribution of nodes' activity potential of resource layer |
| $F_2(y)$ | — | Distribution of nodes' activity potential of epidemic layer |
| P_i^X | $[0, 1]$ | Probability of node i in state X |
| $r_i(t)$ | $[0, 1]$ | Probability that node i has no resources from neighbors at time t |
| $q_i(t)$ | $[0, 1]$ | Probability that node i will not be infected by any neighbors at time t |
| β_c | $[0, 1]$ | The upper threshold of β for the outbreak of epidemic |
| λ_c | $[0, 1]$ | The lower threshold of λ for the outbreak of epidemic |

$\eta_2 \langle y \rangle N$. According to the previous work [49], the activity levels of individuals in social networks usually obey a power-law distribution with a cutoff, and due to the experience from [36], we use the given probability distributions $F_1(x) \propto x^{-\gamma_1}$ and $F_2(y) \propto y^{-\gamma_2}$ to generate the hierarchical activity potential of individuals in the resource and epidemic layer, respectively. Also, we set the lower cutoff $\epsilon = 0.001$ of activity potential, that is, $\epsilon \leq x_i \leq 1$ (y_i is likewise).

Based on the AD temporal two-layer network, we describe the evolving process of coupled RNR-SIS model as the following steps:

- (i) At each time t , node i ($i = 1, \dots, N$) becomes active with probabilities a_i and b_i in the corresponding layer. If the node is active in the resource layer, it will generate m_1 (corresponding m_2 in the epidemic layer) edges to randomly selected nodes. Otherwise, no connections are created. Both active and inactive nodes can receive connections from active nodes.
- (ii) The coupled RNR-SIS dynamic process is run on the constructed two-layer networks. Initially, the nodes only exist in the RS and NI states. The rules are specifically described as follows: (a) only the susceptible nodes can generate resources at the end of

any moment. (b) To some extent, the resourceful nodes lose their all resources with probability δ or diffuse resources to their direct neighbors in the noresource state with probability λ . (c) The susceptible nodes are infected by their direct neighbors in the infected state with probability β . (d) The infected nodes with (or without) resources recover in the susceptible state with probability μ^R (or μ^N). Note that the NS nodes only exist at the transient time but finally disappear at the end of the synchronous-updating coupled dynamic process and the resource diffusion to susceptible nodes is insignificant according to rule (a). Thus, there are only three types of nodes in the coupled dynamic process at each time. Figure 2 shows the transition probability trees of three types of nodes describing the transient changes of their states.

(iii) At next time $t + \Delta t$ (we set $\Delta t = 1$), all edges of the previous network in each layer are deleted and we repeat the construction of AD networks. The coupled dynamic process continuously evolves according to the rules and also is terminated when it converges to the stable state.

3. Theoretical Analysis Based on Microscopic Markov Chain Approach

According to the model mentioned above, each node can be in one of the four states at time t with the probabilities represented by $P_i^{\text{NS}}(t)$, $P_i^{\text{RS}}(t)$, $P_i^{\text{NI}}(t)$, and $P_i^{\text{RI}}(t)$. According to rule (a), we can easily obtain $P_i^{\text{NS}}(t) \equiv 0$. Note that $P_i^{\text{R}} = P_i^{\text{RS}} + P_i^{\text{RI}}$ and $P_i^{\text{I}} = P_i^{\text{NI}} + P_i^{\text{RI}}$. Based on MMCA, we give the following theorems.

Theorem 1. $r_i(t)$ denotes the probability of node i not getting resources from any direct neighbors, which can be written as

$$r_i(t) = \prod_j \left(1 - \lambda \frac{m_1}{N} [a_i P_j^{\text{R}}(t) + a_j P_j^{\text{R}}(t)] \right). \quad (1)$$

The part $(m_1/N)a_i P_j^{\text{R}}(t)$ is the probability that an active node i creates a link with an active node j in the state R and the second part $(m_1/N)a_j P_j^{\text{R}}(t)$ is the probability that an active node i gets a link from an active node j in the state R .

Theorem 2. $q_i(t)$ denotes the probability for the active node i not being infected by any neighbors, which can be written as

$$q_i(t) = \prod_j \left(1 - \beta \frac{m_2}{N} [b_i P_j^{\text{I}}(t) + b_j P_j^{\text{I}}(t)] \right). \quad (2)$$

The part $(m_2/N)b_i P_j^{\text{I}}(t)$ is the probability that an active node i creates a link with an active node j in the state I and the second part $(m_2/N)b_j P_j^{\text{I}}(t)$ is the probability that an active node i gets a link from an active node j in the state I .

Theorem 3. According to Figure 2, we can develop the Markov chains equations for each node i as follows:

$$P_i^{\text{RS}}(t+1) = P_i^{\text{RS}}(t)[\delta q_i(t) + (1-\delta)q_i(t)] + P_i^{\text{NI}}(t)[(1-r_i(t))\mu^R + \mu^N r_i(t)] + P_i^{\text{RI}}(t)[(1-\delta)\mu^R + \delta\mu^N], \quad (3)$$

$$P_i^{\text{NI}}(t+1) = P_i^{\text{RS}}(t)\delta(1-q_i(t)) + P_i^{\text{NI}}(t)(1-\mu^N)r_i(t) + P_i^{\text{RI}}(t)\delta(1-\mu^N), \quad (4)$$

$$P_i^{\text{RI}}(t+1) = P_i^{\text{RS}}(1-\delta)(1-q_i(t)) + P_i^{\text{NI}}(t)(1-r_i(t))(1-\mu^R) + P_i^{\text{RI}}(t)(1-\delta)(1-\mu^R). \quad (5)$$

The theoretical analysis based on MMCA aims to deduce the theoretical outbreak threshold β_c of epidemic spreading through theorems (1)–(3). In the stable state, equations (3)–(5) can satisfy $P_i^{\text{RS}}(t+1) = P_i^{\text{RS}}$, $P_i^{\text{RI}}(t+1) = P_i^{\text{RI}}$, and $P_i^{\text{NI}}(t+1) = P_i^{\text{NI}}$. Near the outbreak threshold of epidemic spreading, the probability of an active node i to be infected is approximately equal to zero, $P_i^{\text{I}} \rightarrow 0$. Therefore, we can obtain

$$q_i \approx 1 - m_2 \beta (b_i \rho^{\text{I}} + \theta_b^{\text{I}}), \quad (6)$$

where $\theta_b^{\text{I}} = \sum_j b_j P_j^{\text{I}}/N$. Next, we give a proof to deduce the theoretical outbreak threshold β_c of epidemic spreading.

Proof. By ignoring the higher-order items, we simplify equations (3)–(5) as

$$P_i^{\text{RS}} = P_i^{\text{RS}}[\delta q_i + (1-\delta)q_i] + P_i^{\text{NI}}[(1-r_i)\mu^R + \mu^N r_i] + P_i^{\text{RI}}[(1-\delta)\mu^R + \delta\mu^N], \quad (7)$$

$$P_i^{\text{NI}} = P_i^{\text{NI}}(1-\mu^N)r_i + P_i^{\text{RI}}\delta(1-\mu^N), \quad (8)$$

$$P_i^{\text{RI}} = P_i^{\text{NI}}(1-\mu^R)(1-r_i) + P_i^{\text{RI}}(1-\delta)(1-\mu^R). \quad (9)$$

It is easy to find that the combination of equations (8) and (9) equals equation (7). By substituting equation (8) into equation (7), consequently, we can rewrite equation (7) as follows:

$$P_i^{\text{RS}}(1-q_i) = \mu^R P_i^{\text{I}} + (\mu^N - \mu^R) \frac{P_i^{\text{NI}}}{1-\mu^N}. \quad (10)$$

Note that $P_i^{\text{RS}} = 1 - P_i^{\text{NI}} - P_i^{\text{RI}}$, and with ignoring higher-order items, the part $P_i^{\text{RS}}(1-q_i)$ can be simplified as $(1 - P_i^{\text{RI}})(1-q_i)$. Then, considering $\xi \rightarrow 0$ (i.e., $\mu^N \ll \mu^R < 1$) and $P_i^{\text{NI}} < P_i^{\text{I}} \ll 1$, we simplify equation (10) as follows:

$$\mu^R P_i^{\text{I}} = (1 - P_i^{\text{RI}})(1 - q_i). \quad (11)$$

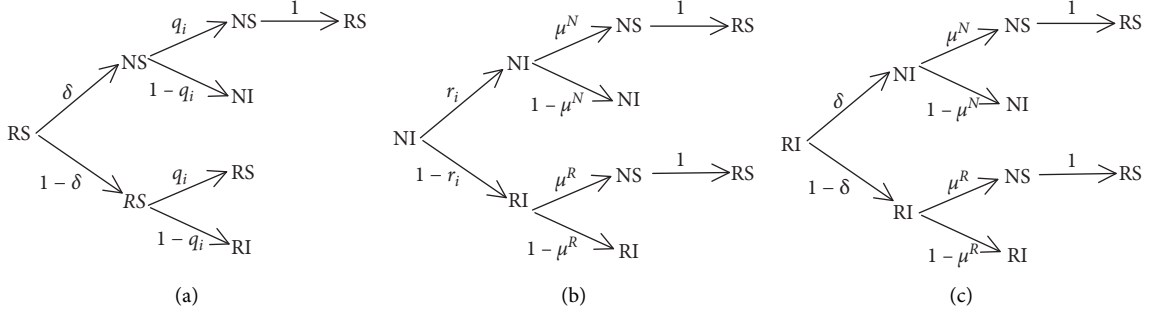


FIGURE 2: Transition probability trees of three types of nodes. They describe the transient changes of nodes' states at each time. The definitions of parameters can be found in Table 1.

Then, we note that $\theta_b^{\text{RI}} = \sum_j b_j^2 P_j^{\text{RI}}/N$. Multiplying b_i on both sides of equation (11) and taking average over all nodes, we obtain a similar expression of θ_b^{I} as

$$\mu^{\text{R}} \theta_b^{\text{I}} = (1 - \theta_b^{\text{RI}})(1 - q_i). \quad (12)$$

Next, we consider equations (11) and (12) as a matrix form

$$\frac{\mu^{\text{R}}}{m_2 \beta} \begin{bmatrix} \rho^{\text{I}} \\ \theta_b^{\text{I}} \end{bmatrix} = \begin{bmatrix} \langle b \rangle - \theta_b^{\text{RI}} & 1 - \rho^{\text{RI}} \\ \langle b^2 \rangle - \theta_b^{\text{RI}} & \langle b \rangle - \theta_b^{\text{RI}} \end{bmatrix} \begin{bmatrix} \rho^{\text{I}} \\ \theta_b^{\text{I}} \end{bmatrix}, \quad (13)$$

β_c is the minimum value of β that satisfies equation (13). And with $\begin{bmatrix} \rho^{\text{I}} \\ \theta_b^{\text{I}} \end{bmatrix} \neq 0$, we can easily solve the critical point with iteration method through the matrix

$$\beta_c = \frac{\mu^{\text{R}}}{m_2} \left\{ \sqrt{(\langle b^2 \rangle - \theta_b^{\text{RI}})(1 - \rho^{\text{RI}}) + \langle b \rangle - \theta_b^{\text{RI}}} \right\}^{-1}. \quad (14)$$

When there is only single layer in which only the epidemic spreads, the outbreak threshold of epidemic spreading could be solved by setting $\theta_b^{\text{RI}} = 0$, $\theta_b^{\text{RI}} = 0$, and $\rho^{\text{RI}} = 0$,

$$\beta_c = \frac{\mu}{m_2 \sqrt{\langle b^2 \rangle + \langle b \rangle}}. \quad (15) \quad \square$$

4. Results

To validate our MMCA based theoretical analysis, we run numerical experiments by the Monte Carlo simulations (MC) and compare the experimental results with theoretical ones regarding different parameters. We construct the coupled network with 1,000 nodes by two AD models. Considering the identical social networks and referring to [36], we set the exponents of probability distributions of activity potential with $\gamma_1 = \gamma_2 = 2.5$, the contact capacities of active nodes with $m_1 = m_2 = 5$, and the rescaling factors with $\eta_1 = \eta_2 = 1000$. The values of the rescaling factor do not affect the following results, but the activity heterogeneity and contact capacity have a significant influence, both of which are verified by the following experimental results. Note that the values of these parameters do not change unless they are chosen as the variables. We represent the final fraction of

infected nodes ρ_{I} and the fraction of noresource nodes ρ_{N} in the stable states, which characterize the coupled dynamic processes of resource diffusion and epidemic spreading. Initially, we set a small fraction of infected nodes with $\rho_{\text{I}} = 0.08$. Meanwhile, to highlight the crucial influence of individual resources on the recovery of disease, we herein set $\xi = 0$ (i.e., $\mu^{\text{N}} = 0$ and $\mu = \mu^{\text{R}}$) that means the infected nodes without resource cannot recover. In the following, we investigate the influences of a number of parameters involved with the RNR-SIS dynamic model and AD network model on the coupled dynamic processes by performing hundreds of MC numerical simulations and obtain a number of numerical results.

Firstly, we investigate the influence of β on the coupled dynamic processes. In Figure 3, it can be found that the changes of ρ_{I} and ρ_{N} with β indicate that there exists a phase transition of ρ_{I} and ρ_{N} when β is larger than the critical points, and the critical points of ρ_{I} and ρ_{N} can be identified by the peak of variability χ calculated as [50]

$$\chi = \omega \frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{\langle \rho \rangle}. \quad (16)$$

Note that, in Figure 3(b), we multiply $\omega = 1000$ to identify the peak more clearly, which does not change the threshold value. For ρ_{I} , its critical point $\beta_c = 0.12$ approximately equals the theoretical outbreak threshold of epidemic spreading by solving equation (14). Moreover, as shown in Figure 3(a), the numerical results of ρ_{I} and ρ_{N} are in good agreement with the theoretical ones obtained by solving equations (1)–(5) (the same in the following figures), which verifies the validation of MMCA based dynamic framework of the coupled model.

Then, we investigate the influences of δ and λ on the coupled dynamic processes, which mainly imply how individual resource diffusion affects the epidemic spreading. Figure 4 shows the changes of ρ_{I} and ρ_{N} with δ and λ , respectively. Specifically, both ρ_{I} and ρ_{N} increase with δ in Figure 4(a) but decrease with λ in Figure 4(b). And, the increasing (or decreasing) trend as a function of δ (or λ) is significantly different between resource diffusion and epidemic spreading. Obviously, δ and λ have a greater influence on resource diffusion than epidemic spreading. More interestingly, a smaller λ (≈ 0.2) can dramatically reduce ρ_{I} ,

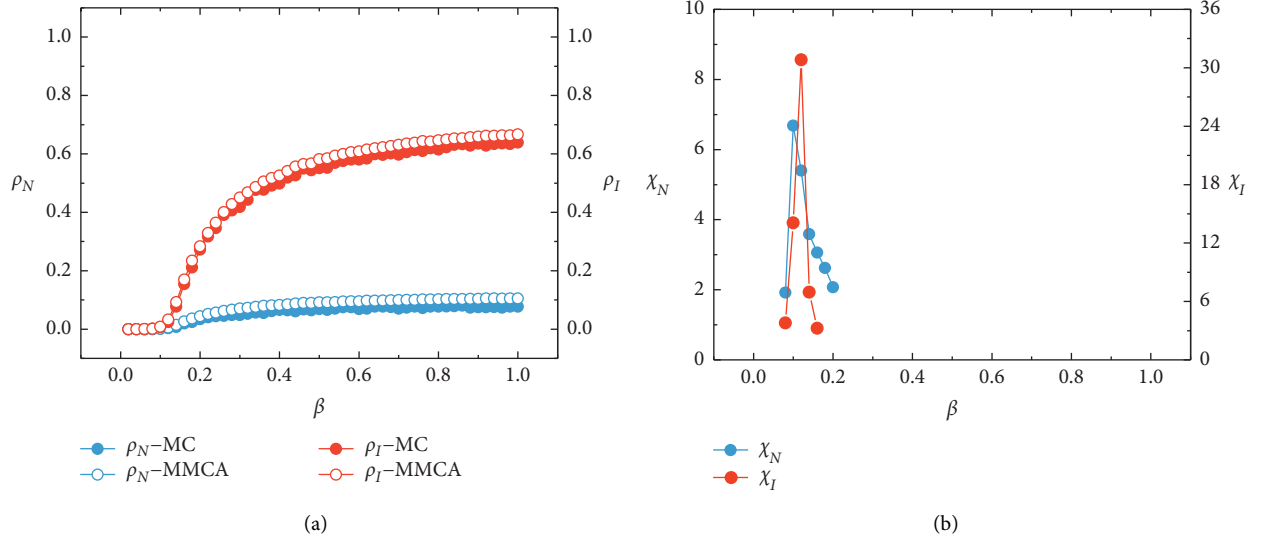


FIGURE 3: (a) The final fractions ρ_I and ρ_N of the infected nodes and the noresource nodes as a function of epidemic transmission rate β . The blue circle indicates ρ_N and the red circle indicates ρ_I . For both ρ_I and ρ_N , we present the results of numerical simulation and MMCA approach in full and open circles, respectively, which are in good agreement with each other. (b) The variability χ of ρ_I and ρ_N as a function of β . We can determine the critical points of ρ_I and ρ_N by the peak of variability and obtain the outbreak threshold of epidemic spreading $\beta_c = 0.12$. The other fixed parameters are $\delta = 0.10$, $\lambda = 0.64$, $\mu = 0.48$.

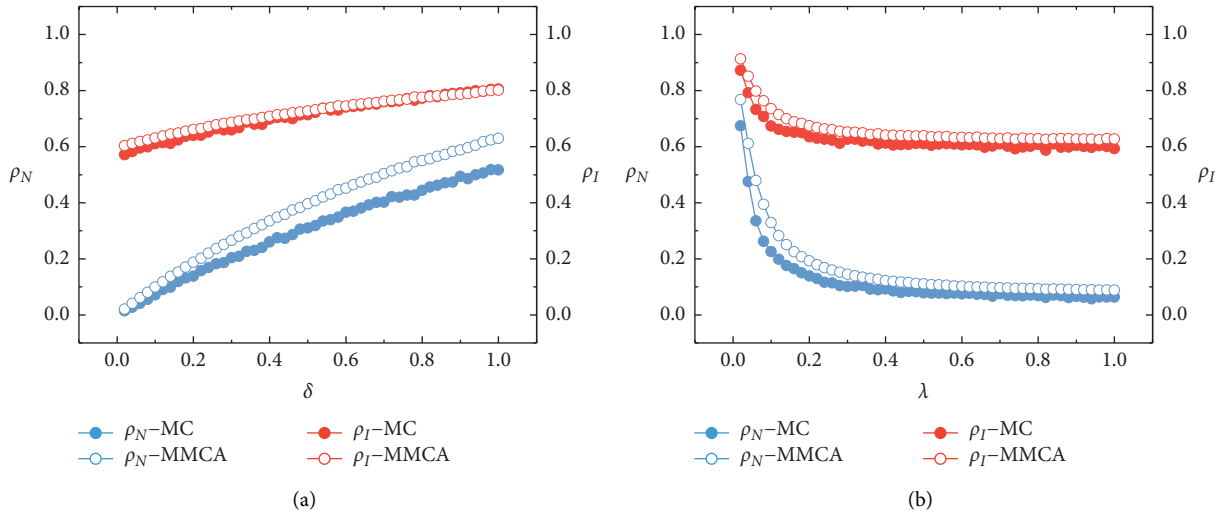


FIGURE 4: The final fractions ρ_I and ρ_N of the infected nodes and the noresource nodes as a function of the resource loss rate δ (a) and the resource diffusion rate λ (b). The symbols are the same as in Figure 3. The results show that both ρ_I and ρ_N increase with δ but decrease with λ . And, a smaller λ (~ 0.2) can dramatically reduce ρ_I , indicating the diffusion of individual resources is of importance for suppressing the spread of disease. The other fixed parameters are $\beta = 0.70$, $\mu = 0.48$.

which, to some extent, suggests the diffusion of individual resources is of importance for suppressing the spread of disease.

We have investigated the influence of single parameter (i.e., δ , λ , and β) on the coupled dynamic process. As the dynamic process of epidemic spreading is more important, we need the comprehensively synergistic influences of these parameters on the epidemic spreading and its outbreak threshold. Figure 5 shows the heat map of ρ_I in respect to a pair of δ (or λ) and β . We can find that, in Figure 5(a), the theoretical outbreak threshold β_c weakly decreases with δ ,

and in Figure 5(b), it quickly increases then barely changes with λ which indicates that the resource diffusion suppresses the outbreak threshold with a lower resource loss rate and larger resource diffusion rate. Meanwhile, when the epidemic outbreaks (i.e., $\beta > \beta_c$), it can be seen that smaller δ and larger λ can more effectively suppress the spread of disease, confirming the conclusion found in Figure 4. We also calculate the (simulated) outbreak threshold of λ obtained by the variability measurement in equation (16) based on the Monte Carlo simulations of numerical experiments, which are shown with the transverse curve in Figure 5(b).

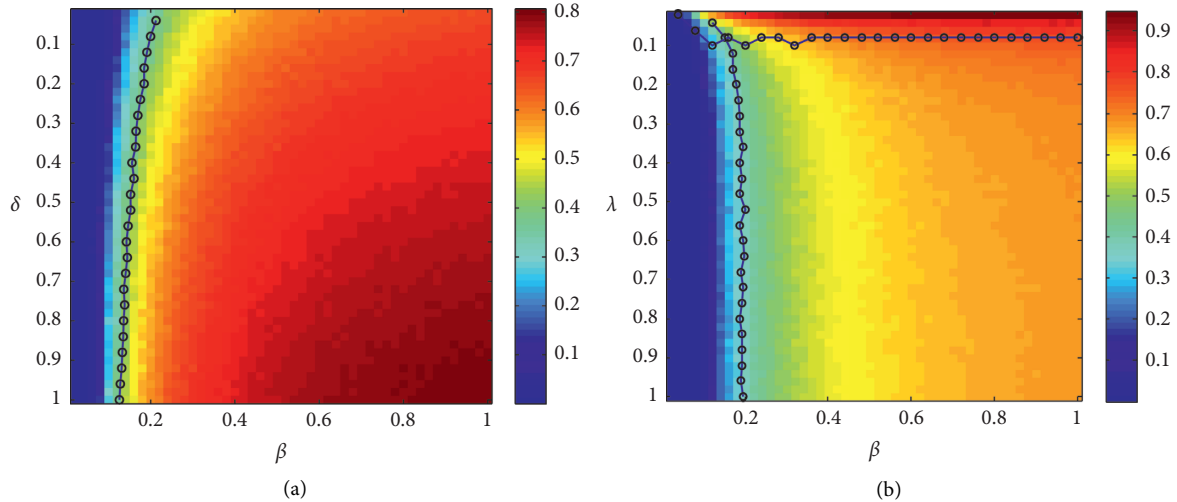


FIGURE 5: The heat maps of ρ_I as a function of a pair of δ and β (a) and λ and β (b). The longitudinal curve in Figures 5(a) and 5(b) indicates the theoretical outbreak thresholds of β obtained by solving equation (13), which is the same as those in Figure 7. And, the transverse curve in Figure 5(b) indicates the simulated outbreak thresholds of λ obtained by the variability measurement, which affects the outbreak of epidemic. It trivially fluctuates only when β is less than 0.2. These two subfigures clearly show that δ and λ affect the outbreak thresholds of each other and effectively affect the outbreak of the epidemic, which suggests that resource diffusion can play an important role in the control of epidemic spreading. The other fixed parameters are $\lambda = 0.64$ for (a), $\delta = 0.10$ for (b), $\mu = 0.48$.

When λ is less than the simulated outbreak threshold, the epidemic will outbreak even if β is very small. We also note that λ_c (i.e., the transverse curve) fluctuates only when β is less than 0.2. A smaller β results in a smaller threshold of resource diffusion rate that a lower rate of epidemic spreading can reduce the risk of an epidemic outbreak caused by a low rate of resource diffusion. However, a smaller λ results in a smaller β_c making it easy for an epidemic outbreak. These results suggest that the diffusive dynamics of individual resources can affect both the outbreak threshold and the spread of disease.

Next, we make efforts to investigate the influences of time-varying network structure mainly including activity heterogeneity and contact capacity of nodes in two AD networks on the coupled dynamic process. Figures 6(a) and 6(b) show that γ_1 and m_1 in the resource layer, respectively, impact ρ_I and ρ_N . They clearly show both ρ_I and ρ_N weekly increase with γ_1 in Figure 6(a) and decrease with m_1 in Figure 6(b). The finding suggests that stronger activity heterogeneity (corresponding to smaller γ_1) and larger contact capacity effectively promote resource diffusion (corresponding to smaller ρ_N) but suppress epidemic spreading (corresponding to smaller ρ_I). Note that, in comparison with the epidemic spreading, γ_1 and m_1 have more influence on the dynamic process of resource diffusion.

In contrast, we present that γ_2 and m_2 in the epidemic layer, respectively, impact ρ_I and ρ_N in Figures 6(c) and 6(d). They clearly shows that both ρ_I and ρ_N weakly decrease with γ_2 in Figure 6(c) and increase with m_2 in Figure 6(d). These results are the opposite to the ones found in Figures 6(a) and 6(b), which suggest that stronger activity heterogeneity

(corresponding to smaller γ_2) and larger contact capacity cause more resource depletion (corresponding to larger ρ_N) and thus promote epidemic spreading (corresponding to larger ρ_I). Also, in comparison with the resource diffusion, we note γ_2 and m_2 having more influences on the dynamic process of epidemic spreading.

The following results also focus on the synergistic influences of γ_i , m_i , and β on the epidemic spreading and its outbreak threshold. It is clearly shown that, in Figure 7(a), the outbreak threshold is almost uncorrelated with γ_1 and, in Figure 7(b), it weakly increases and then barely changes with m_1 . Meanwhile, for $\beta > \beta_c$, we show that the spread of disease is obviously suppressed by more heterogeneous activity potential and larger contact capacity of nodes in the resource layer, which confirms the conclusion found in Figures 6(a) and 6(b). These results indicate that the dynamic social relationships of individuals represented by the AD network model in the resource layer cannot inhibit the outbreak of the disease but can effectively suppress the spread of disease.

In contrast, in Figures 7(c) and 7(d), we can find that the outbreak threshold is strongly affected by γ_2 and m_2 . More specifically, when the activity potentials of individuals in the epidemic layer become more heterogeneous (corresponding to smaller γ_2), the outbreak thresholds dramatically decrease. When the contact capacities of individuals in the epidemic layer are much larger, the outbreak thresholds also dramatically decrease. Meanwhile, for $\beta > \beta_c$, we find that stronger activity heterogeneity and larger contact capacity of individuals in the epidemic layer significantly promote the spread of disease, which also confirms the conclusion found in Figures 6(c) and 6(d). Besides, combining with the results in Figures 7(b) and 7(d), it can be seen that, in respect to each

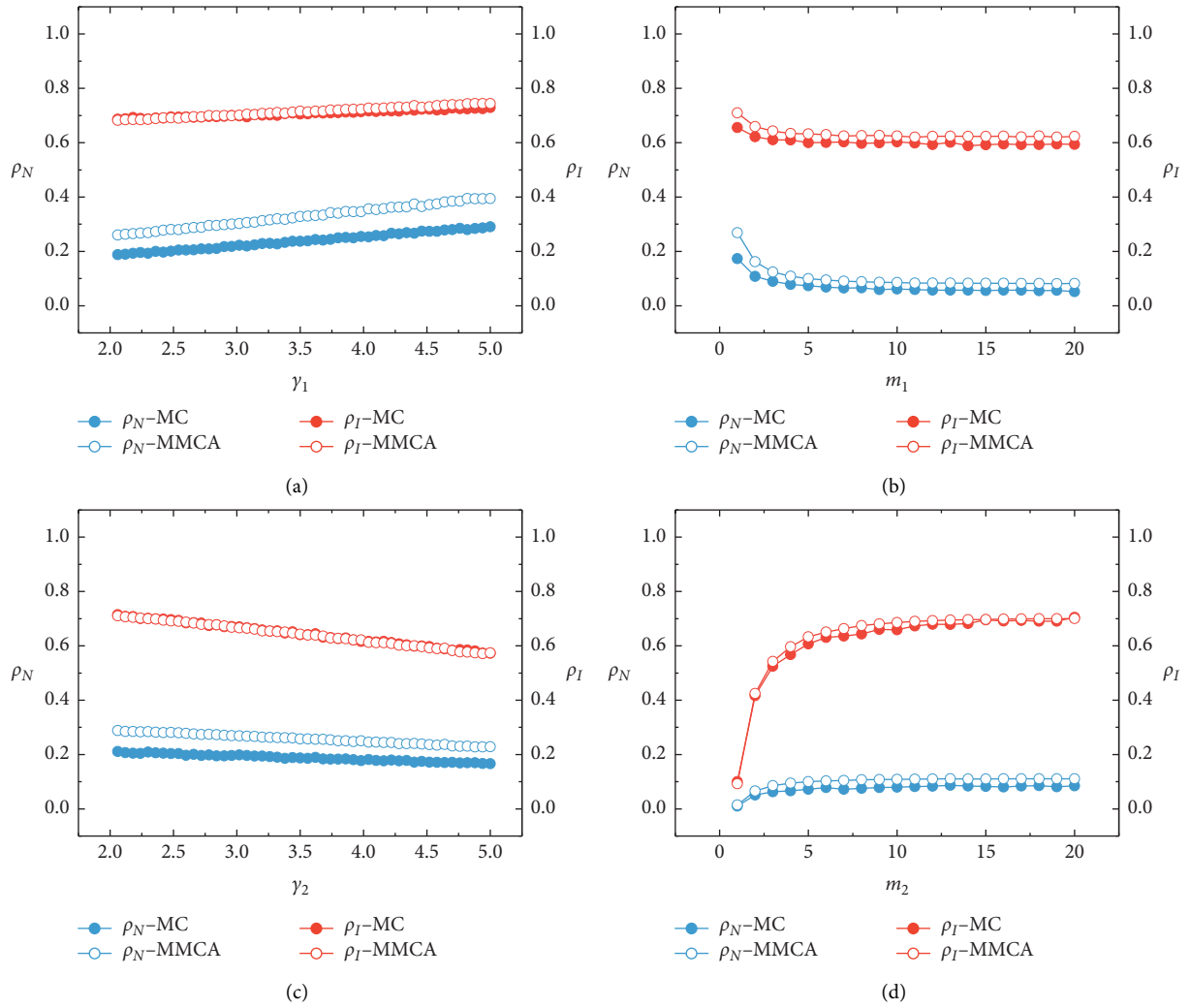


FIGURE 6: The final fractions ρ_I and ρ_N of the infected nodes and the noresource nodes as a function of activity potential γ_i and the contact capacity m_i . The symbols are the same as those in Figure 3. The upper panel presents that ρ_I and ρ_N changes with γ_1 (a) and m_1 (b) in the resource layer, which shows that stronger activity heterogeneity and larger contact capacity of individuals in the resource layer can effectively promote resource diffusion but suppress epidemic spreading. The lower panel presents that ρ_I and ρ_N changes with γ_2 (a) and m_2 (b) in the epidemic layer, which suggests that these two characters of individuals in the epidemic layer cause more resource depletion and thus promote epidemic spreading. The other fixed parameters are $\delta = 0.10$, $\lambda = 0.64$, $\beta = 0.70$, $\mu = 0.48$, and $m_1 = 5$ for (a), (c), and (d); $m_2 = 5$ for (a), (b), and (c); $\gamma_1 = 2.5$ for (b), (c), and (d); and $\gamma_2 = 2.5$ for (a), (b), and (d).

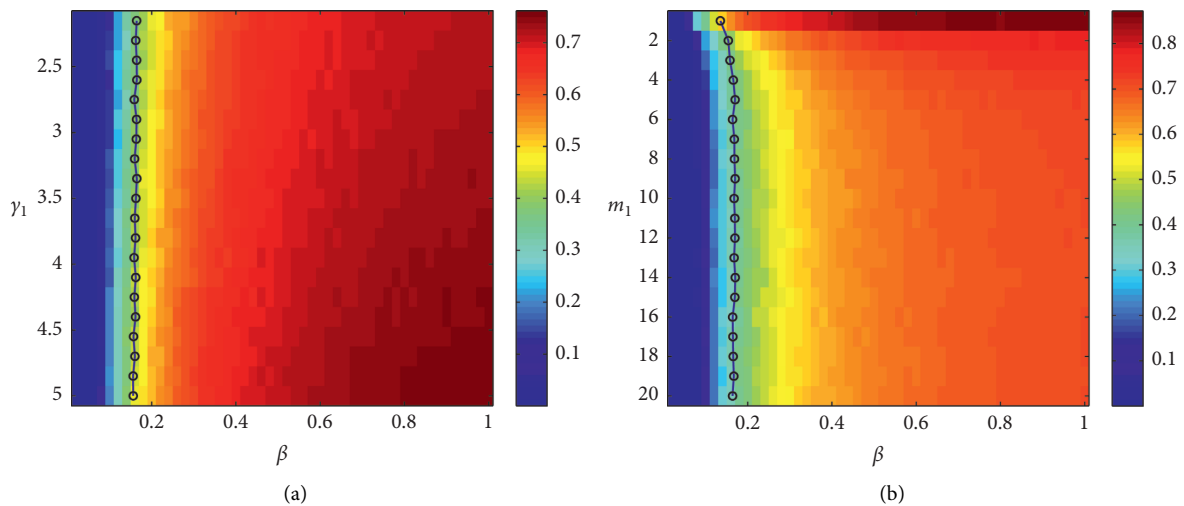


FIGURE 7: Continued.

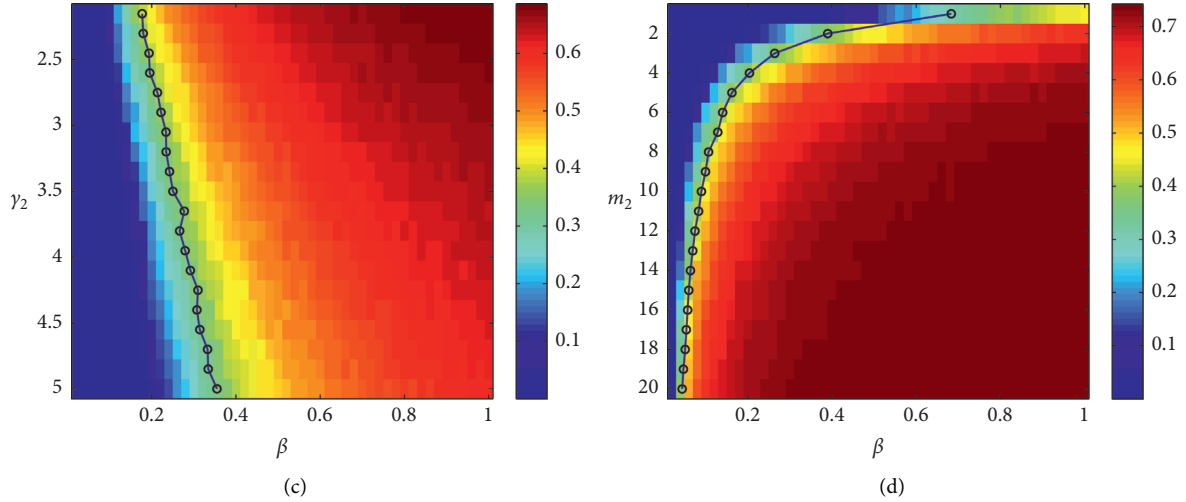


FIGURE 7: The heat maps of ρ_I as a function of a pair of γ_i and β (a–c) and m_i and β (b–d). (a, b) γ_1 and m_1 are approximately uncorrelated with the outbreak threshold and effectively affect the spread of disease, which suggests that the dynamic social relationships of individuals in the resource layer cannot inhibit the outbreak of the disease but can effectively suppress the spreading of disease. (c, d) γ_2 and m_2 significantly affect the outbreak threshold and the spread of disease, which suggests that the dynamic physical contacts of individuals in the epidemic layer have strong influences on the dynamic process of epidemic spreading. The other fixed parameters are $\delta = 0.10$, $\lambda = 0.64$, $\mu = 0.48$, and $m_1 = 5$ for (a), (c), and (d); $m_2 = 5$ for (a), (b), and (c); $\gamma_1 = 2.5$ for (b), (c), and (d); and $\gamma_2 = 2.5$ for (a), (b), and (d).

$\beta (> \beta_c)$, the contact capacity of individuals in both the resource and epidemic layers has a limited influence on the spread of disease.

5. Conclusion

In conclusion, the present work has investigated the coupled dynamic processes of resource diffusion and epidemic spreading via the proposed RNR-SIS dynamic model on time-varying multiplex networks. We perform the investigations by assuming a good abstraction from a real scenario, where the diseases spread on a physical contact network, the individual resources diffuse on an online social network, and they dynamically interact with each other. Considering such two networks are temporal due to the time-varying physical contacts or social relationships of individuals, we first construct the time-varying two-layer network by using two AD network models with different parameters. The AD model encodes the connectivity pattern (i.e., the dynamic process of changing physical contact or social relationship) of individuals in the distribution of activity potential following a power-law function empirically measured in real-world networks. This function allows the definition of a simple dynamic process based on the nodal activity level, providing a time-dependent description of the connectivity pattern.

Then, to explore the full-phase diagram of the coupled RNR-SIS dynamics on the time-varying two-layer network, we develop a probabilistic framework of the intralayer and interlayer dynamic processes of the coupled model by extending MMCA to time-varying multiplex networks. Based on the theoretical analysis of the probabilistic framework, we obtain the expressions of the fractions of infected nodes and noresource nodes and the outbreak

threshold. Monte Carlo simulations are essential to verify the theoretical analysis. Comparing with extensive Monte Carlo simulations of the same system. The approximation using MMCA has an accuracy of up to 1.93% error for the prediction of the epidemic incidence and 21.79% error for the outbreak threshold of β . The principal reason for the large error in β_c is that we neglect some higher-order terms and other related terms. The approximate agreement between theoretical and numerical results indicates the validation of the determined probabilistic framework. Meanwhile, these results show the trivially asymmetric interactions between the coupled dynamic process of resource diffusion and epidemic spreading. That is, the spread of disease is effectively suppressed by the resource diffusion with a lower resource loss rate and larger resource diffusion rate, but the diffusion of individual resources is trivially affected by the epidemic spreading even with a very larger disease transmission rate. Besides, we note that when epidemic outbreaks, the resource diffusion rate has a limited influence on the outbreak threshold and the fraction of infected nodes and also the disease transmission rate affects the threshold of resource diffusion rate limitedly around the threshold of the disease transmission rate.

Next, considering the influence of time-varying two-layer network structure on the coupled dynamics of resource diffusion and epidemic spreading, we mainly focus on how the activity potential and contact capacity of individuals in the resource and epidemic layers affect the final fraction of infected nodes and noresource nodes and the outbreak threshold. The theoretical and numerical results suggest that stronger activity heterogeneity and larger contact capacity of individuals in the resource layer can suppress the spread of disease because they promote the diffusion of individual resources. In contrast, these two characters of individuals in

the epidemic layer tend to mitigate the diffusion of individual resources and thus increase epidemic prevalence. Meanwhile, we note that the outbreak threshold is nearly independent of the time-varying network structure in the resource layer, suggesting the resource diffusion of individuals cannot inhibit the outbreak of disease but only can effectively suppress the spreading of disease. Besides, when the epidemic outbreaks, the contact capacity of individuals in both the resource and epidemic layer has a finite influence on the final fraction of infected nodes, suggesting that increasing contact capacity above a very large value makes no sense at all.

Through the conclusion mentioned above, it can be seen that the diffusion of individual resources interacts closely with the dynamic process of epidemic spreading. On the one hand, the quicker diffusion of individual resources can effectively suppress epidemic spreading. However, only considering the dynamics of resource diffusion, we easily waste the individual resources due to the limited influence of the resource diffusion rate. On the other hand, we turn to analyze how the time-varying network structure affects the coupled dynamic process. When the epidemic outbreaks, individuals with larger activity potential and contact capacity have a high risk of getting infected but also are able to quickly obtain the resources of the neighbors. An extreme scenario is that the lack of individual resources might happen if the epidemic deteriorates, and then it entered a vicious circle. Thus, the obtained conclusion may be useful to provide methods of controlling large-scale epidemic prevalence, such as strengthening resource diffusion, reducing resource loss, speeding up of individuals' activities, and enlarging the contact capacity in social networks. Finally, the proposed coupled resource-epidemic model and analytical approach may be instructive to multiplex spreading models in time-varying multilayer networks.

Data Availability

Correspondence and requests for data should be addressed to S-M. C. (e-mail: shimincai@uestc.edu.cn).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant nos. 61673086, 11975099, and 11575041), the Science and Technology Department of Sichuan Province (Grant no. 2020YFS0007), the Chendu Science and Technology Agency (Grant no. 2020-YF05-00073-SN), and the Science Promotion Programme of UESTC, China (no. Y03111023901014006).

References

- [1] J. Radcliffe and N. T. J. Bailey, "The mathematical theory of infectious diseases and its applications. 2nd edition," *Applied Statistics*, vol. 26, no. 1, pp. 85–87, 1977.
- [2] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction*, Cambridge University Press, Cambridge, UK, 2001.
- [3] R. M. Anderson, B. Anderson, and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, UK, 1991.
- [4] S. Boccaletti, G. Bianconi, R. Criado et al., "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014.
- [5] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of Modern Physics*, vol. 87, no. 3, p. 925, 2015.
- [6] W. Wang, Q.-H. Liu, J. Liang, Y. Hu, and T. Zhou, "Co-evolution spreading in complex networks," *Physics Reports*, vol. 820, pp. 1–51, 2019.
- [7] D. J. Earn, P. Rohani, B. M. Bolker, and B. T. Grenfell, "A simple model for complex dynamical transitions in epidemics," *Science*, vol. 287, no. 5453, pp. 667–670, 2000.
- [8] M. Martcheva and S. S. Pilyugin, "An epidemic model structured by host immunity," *Journal of Biological Systems*, vol. 14, no. 2, pp. 185–203, 2006.
- [9] G. Hartvigsen, J. M. Dresch, A. L. Zielinski, A. J. Macula, and C. C. Leary, "Network structure, and vaccination strategy and effort interact to affect the dynamics of influenza epidemics," *Journal of Theoretical Biology*, vol. 246, no. 2, pp. 205–213, 2007.
- [10] J. O. Hill, J. C. Peters, and H. R. Wyatt, "The role of public policy in treating the epidemic of global obesity," *Clinical Pharmacology & Therapeutics*, vol. 81, no. 5, pp. 772–775, 2007.
- [11] H.-F. Zhang, P.-P. Shu, Z. Wang, M. Tang, and M. Small, "Preferential imitation can invalidate targeted subsidy policies on seasonal-influenza diseases," *Applied Mathematics and Computation*, vol. 294, pp. 332–342, 2017.
- [12] Z. Wang, Y. Moreno, S. Boccaletti, and M. Perc, "Vaccination and epidemics in networked populations-An introduction," *Chaos, Solitons & Fractals*, vol. 103, pp. 177–183, 2017.
- [13] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen, "The spread of awareness and its impact on epidemic outbreaks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6872–6877, 2009.
- [14] S. Samanta, S. Rana, A. Sharma, A. K. Misra, and J. Chattopadhyay, "Effect of awareness programs by media on the epidemic outbreaks: a mathematical model," *Applied Mathematics and Computation*, vol. 219, no. 12, pp. 6965–6977, 2013.
- [15] A. Rizzo, M. Frasca, and M. Porfiri, "Effect of individual behavior on epidemic spreading in activity-driven networks," *Physical Review E*, vol. 90, no. 4, Article ID 042801, 2014.
- [16] K. Paarporn, C. Eksin, J. S. Weitz, and J. S. Shamma, "Networked sis epidemics with awareness," *IEEE Transactions on Computational Social Systems*, vol. 4, no. 3, pp. 93–103, 2017.
- [17] C. Zheng, C. Xia, Q. Guo, and M. Dehmer, "Interplay between sir-based disease spreading and awareness diffusion on multiplex networks," *Journal of Parallel and Distributed Computing*, vol. 115, pp. 20–28, 2018.

- [18] C. Xia, Z. Wang, C. Zheng et al., “A new coupled disease-awareness spreading model with mass media on multiplex networks,” *Information Sciences*, vol. 471, pp. 185–200, 2019.
- [19] H. Yang, C. Gu, M. Tang, S.-M. Cai, and Y.-C. Lai, “Suppression of epidemic spreading in time-varying multiplex networks,” *Applied Mathematical Modelling*, vol. 75, pp. 806–818, 2019.
- [20] M. L. Brandeau, G. S. Zaric, and A. Richter, “Resource allocation for control of infectious diseases in multiple independent populations: beyond cost-effectiveness analysis,” *Journal of Health Economics*, vol. 22, no. 4, pp. 575–598, 2003.
- [21] X. Chen, T. Zhou, L. Feng et al., “Nontrivial resource requirement in the early stage for containment of epidemics,” *Physical Review E*, vol. 100, no. 3, Article ID 032310, 2019.
- [22] N. J. Watkins, C. Nowzari, V. M. Preciado, and G. J. Pappas, “Optimal resource allocation for competitive spreading processes on bilayer networks,” *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 298–307, 2016.
- [23] X.-L. Chen, S.-M. Cai, M. Tang, W. Wang, T. Zhou, and P.-M. Hui, “Controlling epidemic outbreak based on local dynamic infectiousness on complex networks,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 12, Article ID 123105, 2018.
- [24] X. Chen, R. Wang, T. Ming, S. Cai, H. E. Stanley, and L. A. Braunstein, “Suppressing epidemic spreading in multiplex networks with social-support,” *New Journal of Physics*, vol. 20, no. 1, Article ID 013007, 2018.
- [25] X. Chen, W. Wang, S. Cai, H. E. Stanley, and L. A. Braunstein, “Optimal resource diffusion for suppressing disease spreading in multiplex networks,” *Journal of Statistical Mechanics*, vol. 2018, no. 5, Article ID 053501, 2018.
- [26] X.-L. Chen, R.-J. Wang, C. Yang, and S.-M. Cai, “Hybrid resource allocation and its impact on the dynamics of disease spreading,” *Physica A: Statistical Mechanics and Its Applications*, vol. 513, pp. 156–165, 2019.
- [27] R. Wang, X. Chen, and S. Cai, “Local information based resource allocation model for disease suppressing on complex networks,” *Physica A: Statistical Mechanics and Its Applications*, vol. 533, p. 121968, 2019.
- [28] S. Bansal, J. Read, B. Pourbohloul, and L. A. Meyers, “The dynamic nature of contact networks in infectious disease epidemiology,” *Journal of Biological Dynamics*, vol. 4, no. 5, pp. 478–489, 2010.
- [29] E. Valdano, M. R. Fiorentin, C. Poletto, and V. Colizza, “Epidemic threshold in continuous-time evolving networks,” *Physical Review Letters*, vol. 120, no. 6, pp. 068 302.1–068 302.6, 2018.
- [30] J. Sun, L. Feng, J. Xie, X. Ma, D. Wang, and Y. Hu, “Revealing the predictability of intrinsic structure in complex networks,” *Nature Communications*, vol. 11, no. 1, p. 574, 2020.
- [31] R. K. Pan and J. Saramäki, “Path lengths, correlations, and centrality in temporal networks,” *Physical Review E*, vol. 84, Article ID 016105, 2011.
- [32] P.-B. Cui, W. Wang, S.-M. Cai, T. Zhou, and Y.-C. Lai, “Close and ordinary social contacts: how important are they in promoting large-scale contagion?” *Physical Review E*, vol. 98, no. 5, Article ID 052311, 2018.
- [33] X.-H. Chen, S.-M. Cai, W. Wang, M. Tang, and H. E. Stanley, “Predicting epidemic threshold of correlated networks: a comparison of methods,” *Physica A: Statistical Mechanics and Its Applications*, vol. 505, pp. 500–511, 2018.
- [34] P. Block, M. Hoffman, I. J. Raabe et al., “Social network-based distancing strategies to flatten the covid-19 curve in a post-lockdown world,” *Nature Human Behaviour*, vol. 4, pp. 588–596, 2020.
- [35] J. S. Weitz, S. J. Beckett, A. R. Coenen et al., “Modeling shield immunity to reduce covid-19 epidemic spread,” *Nature Medicine*, vol. 26, pp. 849–854, 2020.
- [36] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, “Activity driven modeling of time varying networks,” *Scientific Reports*, vol. 2, p. 469, 2012.
- [37] B. Ribeiro, N. Perra, and A. Baronchelli, “Quantifying the effect of temporal resolution on time-varying networks,” *Scientific Reports*, vol. 3, p. 3006, 2013.
- [38] M.-X. Liu, W. Wang, Y. Liu, M. Tang, S.-M. Cai, and H.-F. Zhang, “Social contagions on time-varying community networks,” *Physical Review E*, vol. 95, no. 5, Article ID 052306, 2017.
- [39] A. Moinet, R. Pastor-Satorras, and A. Barrat, “Effect of risk perception on epidemic spreading in temporal networks,” *Physical Review E*, vol. 97, no. 1, Article ID 012313, 2018.
- [40] M. Newman, “Spread of epidemic disease on networks,” *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 66, Article ID 016128, 08 2002.
- [41] S. Gómez, J. Gómez-Gardenes, Y. Moreno, and A. Arenas, “Nonperturbative heterogeneous mean-field approach to epidemic spreading in complex networks,” *Physical Review E*, vol. 84, no. 3, Article ID 036105, 2011.
- [42] W. Wang, M. Tang, H. E. Stanley, and L. A. Braunstein, “Unification of theoretical approaches for epidemic spreading on complex networks,” *Reports on Progress in Physics*, vol. 80, no. 3, Article ID 036603, 2017.
- [43] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, “Discrete-time Markov chain approach to contact-based disease spreading in complex networks,” *EPL (Europhysics Letters)*, vol. 89, no. 3, p. 38009, 2010.
- [44] C. Granell, S. Gómez, and A. Arenas, “Dynamical interplay between awareness and epidemic spreading in multiplex networks,” *Physical Review Letters*, vol. 111, no. 12, p. 128701, 2013.
- [45] C. Granell, S. Gomez, and A. Arenas, “Competing spreading processes on multiplex networks: awareness and epidemics,” *Physical Review E*, vol. 90, no. 1, Article ID 012808, 2014.
- [46] J.-Q. Kan and H.-F. Zhang, “Effects of awareness diffusion and self-initiated awareness behavior on epidemic spreading—an approach based on multiplex networks,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 44, pp. 193–203, 2017.
- [47] Z. Wang, Q. Guo, S. Sun, and C. Xia, “The impact of awareness diffusion on sir-like epidemics in multiplex networks,” *Applied Mathematics and Computation*, vol. 349, pp. 134–147, 2019.
- [48] Q. Guo, Y. Lei, X. Jiang, Y. Ma, G. Huo, and Z. Zheng, “Epidemic spreading with activity-driven awareness diffusion on multiplex network,” *Chaos*, vol. 26, no. 4, Article ID 043110, 2016.
- [49] Y.-H. Qin, Z.-D. Zhao, S.-M. Cai, L. Gao, and H. E. Stanley, “Dual-induced multifractality in online viewing activity,” *Chaos*, vol. 28, no. 1, Article ID 013114, 2018.
- [50] M. Catanzaro, M. Boguná, and R. Pastor-Satorras, “Generation of uncorrelated random scale-free networks,” *Physical Review E*, vol. 71, no. 2, Article ID 027103, 2005.

Research Article

A New Coupled Awareness-Epidemic Spreading Model with Neighbor Behavior on Multiplex Networks

Chao Zuo, Anjing Wang, Fenping Zhu, Zeyang Meng, and Xueke Zhao 

School of Management Engineering and E-Commerce, Zhejiang Gongshang University, Hangzhou 310018, China

Correspondence should be addressed to Xueke Zhao; zhaoxueke111@gmail.com

Received 11 December 2020; Revised 19 February 2021; Accepted 26 February 2021; Published 18 March 2021

Academic Editor: xiaoke xu

Copyright © 2021 Chao Zuo 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.

In this paper, we propose a nonlinear coupled model to study the two interacting processes of awareness diffusion and epidemic spreading on the same individual who is affected by different neighbor behavior status on multiplex networks. We achieve this topology scenario by two kinds of factors, one is the perception factor that can change interplay between different layers of networks and the other is the neighbors' behavior status that can change the infection rate in each layer. According to the microscopic Markov chain approach (MMCA), we analyze the dynamical evolution of the system and derive the theoretical epidemic threshold on uncorrelated heterogeneous networks, and then, we validate the analysis by numerical simulation and discuss the final size of awareness diffusion and epidemic spreading on a scale-free network. With the outbreak of COVID-19, the spread of epidemic in China prompted drastic measures for transmission containment. We examine the effects of these interventions based on modeling of the awareness-epidemic and the COVID-19 epidemic case. The results further demonstrate that the epidemic spreading can be affected by the effective transmission rate of the awareness and neighbors' behavior status.

1. Introduction

The outbreak of COVID-19 can involve the diffusion of information about the epidemic, including the officially released authoritative information, rumors, and fears [1–3]. In reality, awareness diffusion can stimulate individuals to take spontaneous behavioral responses such as wearing masks or staying at home to reduce the frequency of face-to-face contact [4, 5]. Therefore, the epidemic-spreading dynamics in complex networks has attracted increasing attention in many disciplines [6, 7]. However, the exact impact they can have on the epidemic dynamics is difficult to quantify, so mathematical modelling is used to test hypotheses and identify pivotal parameters in the interplay between awareness diffusion and epidemic spreading [8–11]. For example, Granell et al. constructed a UAU-SIS (unaware-aware-unaware/susceptible-infected-susceptible) model to study the interaction between epidemics spreading and awareness diffusion on multiplex networks [12]. Funk et al. researched the coevolution of information and epidemic on public and found that the spread of

positively oriented information about the epidemic can suppress the epidemic [13]. Wang et al. utilized real data to investigate the coevolution mechanisms between information and disease spreading, and the empirical analysis showed that there is an asymmetric interaction between information and disease spreading [14]. Moreover, Wang et al. further found that the awareness inhibits the epidemic spreading, whereas the epidemic spreading facilitates the awareness diffusion [15]. Zhu et al. thought that an epidemic might spread among multicommunities, so they modeled each community as a multiplex network that included both a virtual awareness layer and a physical layer [16]. They further presented a coupled UAU-SIRD model, which consisted of a virtual layer sustaining unaware-aware-unaware dynamics and a physical layer supporting the susceptible-infected-recovered-dead process, to investigate the spreading property of epidemics and the relationships between the focused variables and parameters of the epidemic. The result indicated that the incorporation of virtual layers would reduce the range of affected individuals [17]. Kabir and Tanimoto established

a two-layer SIR-UA (susceptible-infected-recovered/unaware-aware) epidemic transmission model that comprehensively considered the influence of individual behavior and awareness diffusion on epidemic spreading on heterogeneous networks [18]. Shang studied the effects of global, local, and contact awareness on a discrete-time SIS epidemic dynamics and used the stability theory of matrix difference equation to derive the epidemic threshold [19]. They further investigated the impact of the three forms of awareness on epidemic spreading by the mean-field approach with heterogeneous transmission rates, and the numerical simulation results showed that both local and contact awareness can raise the epidemic threshold while global awareness can only decrease the final epidemic size [20].

Though the effects of information-based behavioral responses on the epidemic dynamics have been studied by many documents, most of them assumed that individuals are treated equally when they contact different neighbors. In fact, the deviation among different neighbors has a remarkable influence on the individual infection process due to the complex topological structures of networks. In view of this, there has been an increasing focus on the discussion of neighbor behavior in the interacting processes of awareness diffusion and epidemic spreading [21–25]. For instance, Kan and Zhang assumed that susceptible individuals can not only be informed by aware neighbors in the information network but also by self-awareness induced by the infected neighbors in the contact network. Results indicated that the individuals with more neighbors in the information network have higher awareness acquisition and they are hard to be infected in the contact network [26]. Guo et al. introduced a heterogeneous threshold model to converse that a hub with a large number of neighbors is relatively easy to become infected [6]. Moreover, they proposed a heterogeneous spreading model that considers the degree heterogeneity and k -core measures heterogeneity of individuals with the belief that different individuals facing the same epidemic would exhibit distinct behaviors according to their own experiences and attributes [27]. Zhu et al. generated an adaptive strategy of the susceptible/exposed individual that a susceptible/exposed individual would disconnect with infected neighbors probabilistically and one new connection would be constructed [28]. Chen et al. proposed a resource-epidemic coevolution model to investigate the effects of the heterogeneous distribution of self-awareness and the heterogeneous distribution of node degree on the epidemic dynamics. They found that the heterogeneity of self-awareness distribution suppresses the outbreak of an epidemic and the heterogeneity of degree distribution enhances the epidemic spreading [29]. Pan and Yan proposed a coupled awareness-epidemic spreading model incorporating the heterogeneity of neighbors' responses to disease outbreaks, and the result showed that the heterogeneity of neighbors' responses acts on the epidemic threshold in the higher stage [30].

Some research studies assumed that the diffusion of awareness will reduce the infection rate of individuals [12–16]. Considering that the spread rate of awareness is also affected

by the spread of the epidemic [15], in this work, the awareness perception factor f_1 ($0 \leq f_1 \leq 2$) stands for the influence of awareness diffusion on the spread rate of epidemic, and we also introduce an epidemic perception factor f_2 ($0 \leq f_2 \leq 2$) to express the impact of epidemic spreading on the diffusion rate of awareness. We define $f_1 \leq 1$, $f_2 > 1$ when authoritative information and epidemic interact and rule $f_1 > 1$, $f_2 \leq 1$ when rumors and epidemic interact, as shown in Table 1.

Moreover, Kan and Zhang discussed the effect of infected neighbors on the formation of node self-awareness [26], and some references confirmed that the heterogeneity of degree distribution enhances the epidemic spreading [27, 29]. On this basis, we assume that the health of awareness neighbors (susceptible, infected, and recovered) plays different roles in the formation of node self-awareness and propose the health-impact-awareness status factors ($\Delta\rho_1$, $\Delta\rho_2$, and $\Delta\rho_3$) to distinguish this effect in the awareness layer; the heterogeneity of neighbors degree distribution will affect the individual's epidemic infection rate, and the behavior-impact-epidemic status factor ($1 - e^{(-k_j/\sum k)}$, a positive correlation function of neighbors degree) is established to discriminate the difference in the epidemic layer. Therefore, we propose a multiplex networks model to comprehend the spreading dynamics between epidemic and awareness on the same population with the influence of different neighbor behavior status. We introduce the behavior status to each node in the multiplex networks. On the one hand, people can acquire awareness from aware neighbors with different health-impact-awareness status in layer 1 of fictitious contacts, such as Microblogs, WeChat, or other social media. On the other hand, in layer 2 of physical contacts, the level of epidemic spreading would be influenced by the infected neighbors with different behavior-impact-epidemic status. The MMCA theoretical analysis and numerical simulations results reveal that awareness and behavioral changes could have favorable effects on the epidemic spreading.

The rest of this paper is organized as follows: in Section 2, we introduce the multiplex networks model. Then, we analyze the dynamical processes of the model with the MMCA method and derive the expression of epidemic threshold in Section 3. Next, in Sections 4 and 5, we perform numerical simulations and discuss the Chinese COVID-19 epidemic case to validate theoretical predictions. Finally, we summarize our findings and conclusions in Section 6.

2. Nonlinear Coupled Awareness-Epidemic Model

2.1. Model Descriptions. In this work, we generalize a multiplex networks model to formalize and simplify the spread mechanisms, as illustrated in Figure 1. In layer 1, people share epidemic-prevention awareness through social media, while people get infected in layer 2. Neighbor behavior is simplified as the “health-impact-awareness” status of aware neighbors in layer 1 and the “behavior-impact-epidemic” status of infected neighbors in layer 2. The definitions of key parameters are shown in Table 2. We suppose the multiplex propagation process according to the following regulations:

TABLE 1: Parameter meaning and influence on individual behavior.

| Parameter | Model | Individual behavior |
|-----------|--|--|
| $f_1 < 1$ | Authoritative information and epidemic | The ability of conscious individuals to prevent virus increases as f_1 decreases |
| $f_1 > 1$ | Rumors and epidemic | The contribution of conscious individuals to virus spread increases as f_1 increases |
| $f_2 > 1$ | Authoritative information and epidemic | The willingness of infected individuals to transmit information increases as f_2 increases |
| $f_2 < 1$ | Rumors and epidemic | The resistance of infected individuals to rumor increases as f_2 decreases |

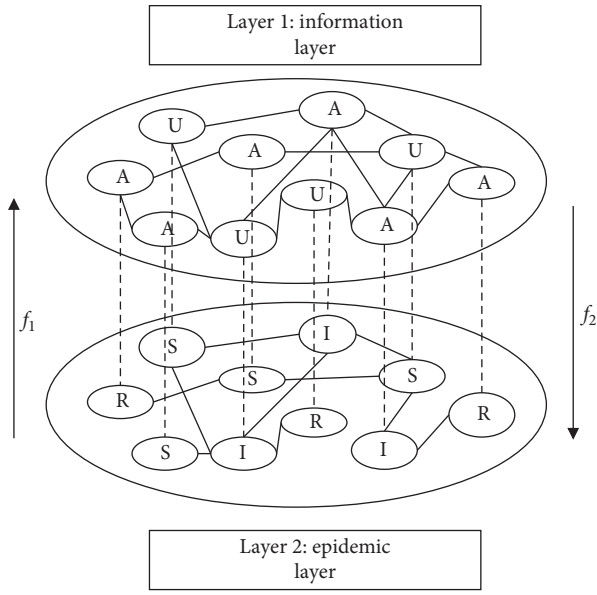


FIGURE 1: UAU-SIR model. Awareness is disseminating in layer 1, and the nodes have two possible states: unaware (U) or aware (A). Epidemic is spreading in layer 2, where the nodes have three possible states: susceptible (S), infected (I), and recovered (R).

- (i) Awareness diffusion in layer 1: the diffusion of awareness satisfies the UAU (unaware-aware-unaware) process. Aware (A) individuals can apply their awareness to decrease the probability of being infected. Unaware (U) individuals do not have any awareness about the epidemic prevention. Unaware individuals can contact with aware neighbors to get awareness with probability λ , while the aware individuals can forget the awareness with probability δ . As we all know, in many cases, individuals who are infected or cured are more aware than those who are far away from the epidemic because once they are exposed to a certain infectious disease (such as SARS or COVID-19), they will be isolated and feel panicked, that is, whether an individual has been physically exposed to the epidemic changes the probability of awareness acquisition. Therefore, we assume that unaware infected (UI) and unaware recovered (UR) individuals know about the epidemic and can increase the probability of being aware with the epidemic perception factor f_2 . Then,

the awareness infection rate can be represented by the susceptible infection rate $\lambda_S = \lambda$ and infected (recovered) infection rate $\lambda_I = f_2 \lambda$ ($\lambda_R = f_2 \lambda$), respectively.

- (ii) Epidemic spreading in layer 2: the spreading of epidemic satisfies the SIR (susceptible-infected-recovered) process. Susceptible individuals get infected with the probability β , while the infected individuals recover with the probability μ . Yet, aware susceptible (AS) individuals can apply the awareness to decrease the probability of being infected with the awareness perception factor f_1 . Then, the epidemic infection rate can be expressed by the unawareness of infection rate $\beta_U = \beta$ and awareness of infection rate $\beta_A = f_1 \beta$, respectively.
- (iii) The health-impact-awareness status of aware neighbors: the direct application of the UAU model in the two-layer network means that the default importance ratio of aware neighbors is AS:AI:AR=1/3:1/3:1/3. We believe that the aware neighbors of different health states play different roles in the process of awareness diffusion and assume that the importance ratio is AS:AI:AR=1/3 + $\Delta\rho_1$:1/3 + $\Delta\rho_2$:1/3 + $\Delta\rho_3$ ($\Delta\rho_1 + \Delta\rho_2 + \Delta\rho_3 = 0$). $\Delta\rho_1$, $\Delta\rho_2$, and $\Delta\rho_3$ are the health-impact-awareness status factors of the model, as shown in Figure 2.
- (iv) The behavior-impact-epidemic status of infected neighbors: the heterogeneity of degree distribution plays a role in the epidemic spreading, and we believe that infected individuals with a large range of daily activities (a large node degree) are more likely to carry and transmit the virus. We use the degree k_j of neighbor j to represent its daily activity range and use $\sum k$ to represent the total degrees of all infected neighbors; as shown in Figure 3, $1 - e^{-(k_j/\sum k)}$ can distinguish the impact of each infected neighbor on epidemic infected probability and also limit the size of this impact (the interval is 0-0.5).

2.2. *Dynamic Model.* According to these presumptions, there are six primary states: US (unawareness susceptibility); UI (unawareness infection); UR (unawareness recovered); AS (awareness susceptibility); AI (awareness infected); and

TABLE 2: The definitions of key parameters.

| Parameter | Description |
|--|--|
| β | Probability of getting infected for susceptible individuals (the basic infection rate) |
| μ | Probability of recovery |
| λ | Probability of becoming aware (the basic infection rate) |
| δ | Probability of becoming unaware |
| f_1 | The awareness perception factor |
| f_2 | The epidemic perception factor |
| $\Delta\rho_1, \Delta\rho_2,$ and $\Delta\rho_3$ | The health-impact-awareness status influence of aware neighbors |
| $(1 - e^{-(k_j/\sum k)})$ | The behavior-impact-epidemic status influence of infected neighbors |

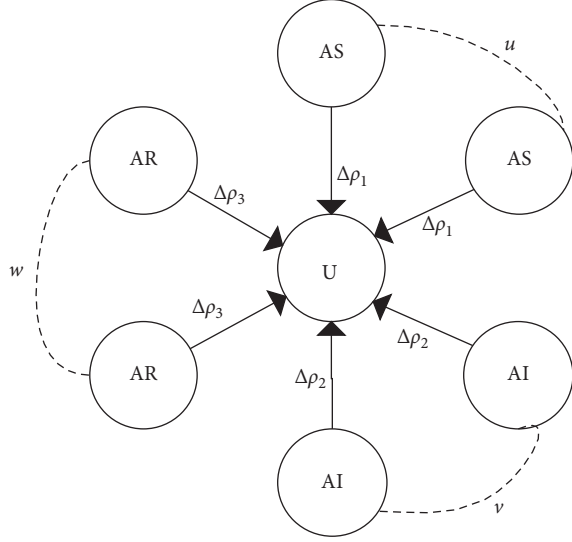


FIGURE 2: Explanation of the health-impact-awareness status of aware neighbors in layer 1. The numbers of AS, AI, and AR are u , v , and w , respectively. Also, the health-impact-awareness factors of the three aware neighbors are $\Delta\rho_1$, $\Delta\rho_2$, and $\Delta\rho_3$, respectively.

AR (awareness recovered). The reaction procedure of the model can be schematically expressed as follows:

- (i) Awareness diffusion $US + AS \xrightarrow{\lambda, 1 + \Delta\rho_1} AS + AS$, $US + AI \xrightarrow{\lambda, 1 + \Delta\rho_2} AS + AI$, $US + AR \xrightarrow{\lambda, 1 + \Delta\rho_3} AS + AR$, $UI + AS \xrightarrow{f_2 * \lambda, 1 + \Delta\rho_1} AI + AS$, $UI + AI \xrightarrow{f_2 * \lambda, 1 + \Delta\rho_2} AI + AI$, $UI + AR \xrightarrow{f_2 * \lambda, 1 + \Delta\rho_3} AI + AR$, $UR + AS \xrightarrow{f_2 * \lambda, 1 + \Delta\rho_1} AR + AS$, $UR + AI \xrightarrow{f_2 * \lambda, 1 + \Delta\rho_2} AR + AI$, $UR + AR \xrightarrow{f_2 * \lambda, 1 + \Delta\rho_3} AR + AR$
- (ii) Epidemic spreading $US + I \xrightarrow{\beta, 1 + (1 - e^{-(k_j/\sum k)})}$, $AI + I, AS + I \xrightarrow{f_1 * \beta, 1 + (1 - e^{-(k_j/\sum k)})}$, $AI + I \xrightarrow{1 + (1 - e^{-(k_j/\sum k)})} AI + I$

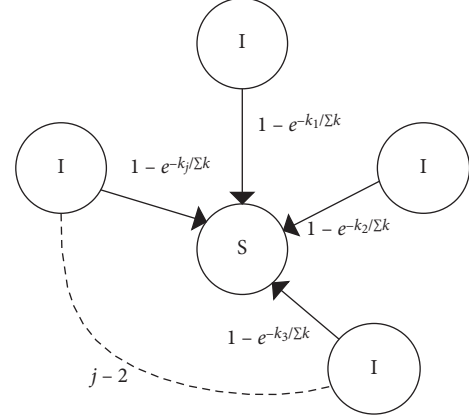


FIGURE 3: Illustration of the behavior-impact-epidemic status of infected neighbors in layer 2. j is the total number of an individual's infected neighbors, and $1 - e^{-(k_j/\sum k)}$ is a positive correlation function of each neighbor's degree.

- (iii) Recoveries $I \xrightarrow{\mu} R$, $A \xrightarrow{\delta} S$

3. Microscopic Markov Chain Approach

We analyze the dynamical processes of the model with the MMCA method [31, 32]. Let x_{ij} and y_{ij} be the adjacency matrices that support the UAU and SIR processes, respectively. The probability of node i in one of the six states at time t is denoted by $P_i^{US}(t)$, $P_i^{UI}(t)$, $P_i^{UR}(t)$, $P_i^{AS}(t)$, $P_i^{AI}(t)$, and $P_i^{AR}(t)$, respectively. Provisions $P_i^{US}(t) + P_i^{UI}(t) + P_i^{UR}(t) + P_i^{AS}(t) + P_i^{AI}(t) + P_i^{AR}(t) = 1$. Under the assumption that the possibilities of becoming aware or infected by any neighbor are independent, we use $r_i^*(t)$, $r_i'(t)$, q_i^{A*} , and q_i^{U*} to represent four basic quantities as follows:

$$\begin{aligned}
 r_i^*(t) &= \Pi_u \left\{ \left[1 - x_{ui} P_u^{AS}(t) \lambda \right]^{(1+\rho_1)} \right\} + \Pi_v \left\{ \left[1 - x_{vi} P_v^{AI}(t) \lambda \right]^{(1+\rho_2)} \right\} + \Pi_w \left\{ \left[1 - x_{wi} P_w^{AR}(t) \lambda \right]^{(1+\rho_3)} \right\}, \\
 r_i'(t) &= \Pi_u \left\{ \left[1 - x_{ui} P_u^{AS}(t) f_2 \lambda \right]^{(1+\rho_1)} \right\} + \Pi_v \left\{ \left[1 - x_{vi} P_v^{AI}(t) f_2 \lambda \right]^{(1+\rho_2)} \right\} + \Pi_w \left\{ \left[1 - x_{wi} P_w^{AR}(t) f_2 \lambda \right]^{(1+\rho_3)} \right\},
 \end{aligned} \tag{1}$$

where $r_i^*(t)$ denotes the probability for susceptible node i not getting aware by any neighbors. $r'_i(t)$ is the probability for infected (recovered) node i not getting aware by any neighbors. x_{ui} means that node i has u neighbors in state P_i^{AS} :

$$\begin{aligned} q_i^{A*} &= \Pi_j [1 - y_{ji} P_j^I(t) f_1 \beta]^{1 + (1 - e^{-(k_j/\sum k)})}, \\ q_i^{U*} &= \Pi_j [1 - y_{ji} P_j^I(t) \beta]^{1 + (1 - e^{-(k_j/\sum k)})}, \end{aligned} \quad (2)$$

where q_i^{A*} (q_i^{U*}) denotes the probability of susceptible aware (unaware) node i not being infected by any neighbors and $y_{ji} = 1$ means that node j is the neighbor of node i at the same layer. We extend Markov's theory to steady state and

assume that $P_i^{AI} + P_i^{UI} = P_i^I = \pi_i \ll 1$. We can obtain the approximations of q_i^{A*} and q_i^{U*} as follows:

$$\begin{aligned} q_i^{A*} &\approx \left(1 - f_1 \beta \sum_j \left(2 - e^{-(k_j/\sum k)} \right) y_{ji} \pi_j \right), \\ q_i^{U*} &\approx \left(1 - \beta \sum_j \left(2 - e^{-(k_j/\sum k)} \right) y_{ji} \pi_j \right). \end{aligned} \quad (3)$$

The potential state transition process of the UAU-SIR model is shown in Figure 4.

As shown in Figure 4, every time step is divided into six stages, and we can easily get the MMCA equations for node i as follows:

$$\begin{cases} P_i^{US}(t+1) = P_i^{US}(t) r_i^*(t) q_i^{U*}(t) + P_i^{AS}(t) \sigma q_i^{U*}(t), \\ P_i^{UI}(t+1) = P_i^{US}(t) r_i^*(t) [1 - q_i^{U*}(t)] + P_i^{UI}(t) r'_i(t) (1 - \mu) + P_i^{AS}(t) \sigma [1 - q_i^{U*}(t)] + P_i^{AI}(t) \sigma (1 - \mu), \\ P_i^{UR}(t+1) = P_i^{UI}(t) r'_i(t) \mu + P_i^{UR}(t) r'_i(t) + P_i^{AI}(t) \sigma \mu + P_i^{AR}(t) \sigma, \\ P_i^{AS}(t+1) = P_i^{US}(t) (1 - r_i^*(t)) q_i^{A*}(t) + P_i^{AS}(t) (1 - \sigma) q_i^{A*}(t), \\ P_i^{AI}(t+1) = P_i^{US}(t) (1 - r_i^*(t)) [1 - q_i^{A*}(t)] + P_i^{UI}(t) [1 - r'_i(t)] (1 - \mu) + P_i^{AS}(t) (1 - \sigma) [1 - q_i^{A*}(t)] + P_i^{AI}(t) (1 - \mu) (1 - \sigma), \\ P_i^{AR}(t+1) = P_i^{UI}(t) (1 - r'_i(t)) \mu + P_i^{UR}(t) (1 - r'_i(t)) + P_i^{AI}(t) (1 - \sigma) \mu + P_i^{AR}(t) (1 - \sigma). \end{cases} \quad (4)$$

The solution of equation (4) is a set of fixed-point equations that satisfy $P_i^{AI}(t+1) = P_i^{AI}(t) = P_i^{AI}$, and this relationship is also true for the other five types of nodes. We add the second and the fifth equations in equation (4) as follows:

$$\begin{aligned} \mu P_i^I &= P_i^{US}(t) \{ r_i^*(t) [1 - q_i^{U*}(t)] + (1 - r_i^*(t)) [1 - q_i^{A*}(t)] \} \\ &\quad + P_i^{AS}(t) \{ \sigma [1 - q_i^{U*}(t)] + (1 - \sigma) [1 - q_i^{A*}(t)] \}. \end{aligned} \quad (5)$$

Near the threshold, the probability of nodes being infected is very small, i.e., $P_i^{AI} + P_i^{UI} = P_i^I = \pi_i \ll 1$. Taking equation (3) into equation (4) and omitting the $O(\epsilon)$ terms, we get

$$\begin{cases} P_i^{US} = P_i^{US} r_i^*(t) + P_i^{AS}(t) \sigma, \\ P_i^{AS} = P_i^{US} (1 - r_i^*(t)) + P_i^{AS} (1 - \sigma). \end{cases} \quad (6)$$

Substituting (6) into (5), we get

$$\begin{aligned} \mu P_i^I &= \mu \pi_i = P_i^{US} [1 - q_i^{U*}(t)] + P_i^{AS} [1 - q_i^{A*}(t)] \\ &= (P_i^{US} + f_1 P_i^{AS}) \beta \sum_j \left(2 - e^{-(k_j/\sum K)} \right) y_{ji} \pi_j, \end{aligned} \quad (7)$$

with $P_i^{AS} \approx P_i^A$, $P_i^{US} \approx 1 - P_i^A$, then, equation (7) can be rewritten as

$$\sum_j \left\{ \left[1 - (1 - f_1) P_i^A \right] \left(2 - e^{-(k_j/\sum k)} \right) y_{ji} - \frac{\mu}{\beta} \tau_{ij} \right\} = 0, \quad (8)$$

where τ_{ij} denotes the elements of the identity matrix. We define a new matrix H , where $h_{ij} = [1 - (1 - f_1) P_i^A] (2 - e^{-(k_j/\sum k)}) y_{ji}$, to simplify equation (8). The epidemic threshold is equal to the minimum value of β that satisfies equation (8). We can obtain the epidemic threshold by denoting $\Lambda_{\max}(h)$, the maximum eigenvalue of H , and the threshold formula is

$$\beta_c = \frac{\mu}{\Lambda_{\max}(H)}. \quad (9)$$

Equations (8) and (9) show that the epidemic threshold is based on the structure of layer 2 (y_{ji}), the parameter f_1 , and the density of awareness P_i^A . The value of P_i^A is further determined by the structure of layer 1 (x_{ji}), the transmission rate λ , and the recovery rate σ .

4. Results

We discuss the effect of f_1 and f_2 through Model I, on which we introduce the neighbor behavior state, and obtain Model II. The comparison between Model 1 and Model 2 reflects the influence of the neighbor behavior state on the propagation process. We perform extensive Monte Carlo numerical simulations [33, 34] for the model (run 100 times) and obtain the infection density ρ_I and aware density ρ_A at steady state. $\rho_I = (\sum_i \rho_i^I / N) = (\sum_i \rho_i^{AI} / N)$ and $\rho_A = (\sum_i \rho_i^A / N)$, where N represents the number of all individuals in the model. In our simulations, the size of the two-layer network is $N = 1000$, the initial number of nodes $m_0 = 5$, the number of connected edges $m = 3$, and the average degree $\langle k \rangle = 6$. The default value of the parameter is

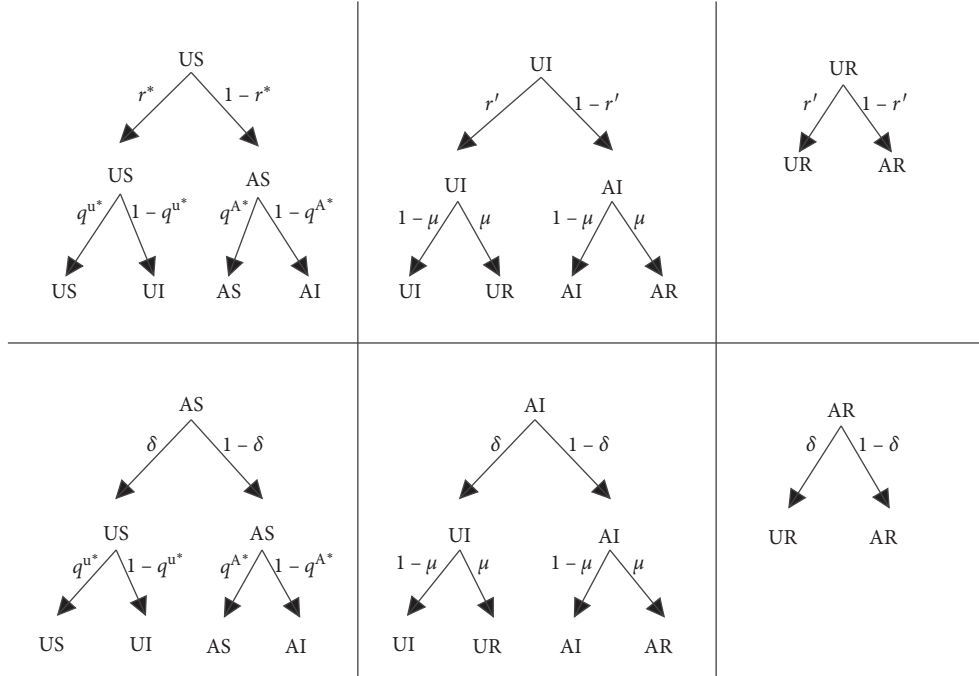
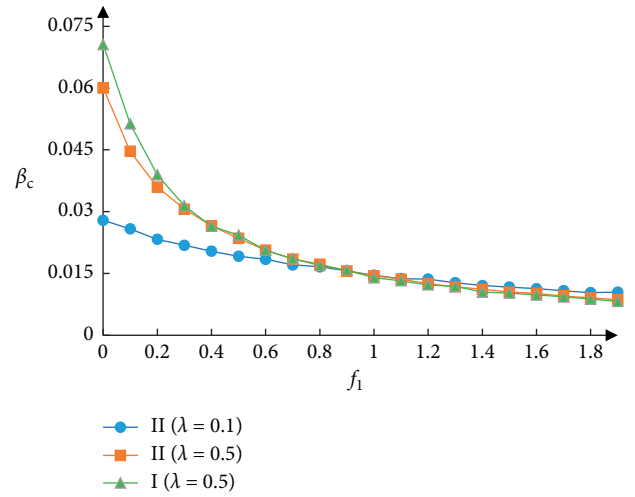


FIGURE 4: Transition probability trees for six possible states.

fixed as $\lambda = 0.5, \sigma = 0.3, \beta = 0.5, \mu = 0.2, f_1 = 1, f_2 = 1, \Delta\rho_1 = 0.067 - 1/3, \Delta\rho_2 = 0.689 - 1/3, \Delta\rho_3 = 0.244 - 1/3$ (the values of $\Delta\rho_1, \Delta\rho_2$, and $\Delta\rho_3$ in this paper can be set according to actual requirements. We believe that the awareness-spreading intensity of different neighbors is $AI > AR > AS$, and by the summation product method and consistency test, we obtain the importance ratio $AS : AI : AR = 0.067, 0.689, 0.244$).

4.1. Threshold. Figure 5 plots the density of β_c as a decreasing function of f_1 . It is clear that β_c obviously decreases as f_1 ($f_1 \in (0, 1)$) increases and basically remains unchanged as f_1 , ($f_1 \in (1, 2)$) increases. Figure 6 presents that β_c grows as λ increases ($f_1 = 0.2, f_2 = 1.8$) and β_c decreases as λ increases ($f_1 = 1.8, f_2 = 0.2$), which means authoritative information can inhibit the outbreak of the epidemic and rumors will contribute to the spread of the epidemic. At the same time, a smaller information forgetting rate, a greater rumor forgetting rate, and a greater infectious diseases recovery rate can raise the epidemic threshold. Moreover, the curve of Model II is lower than that of Model I with equal parameters in both Figures 5 and 6; measures such as locating and isolating infected neighbors are good ways to prevent an outbreak of the epidemic.

As shown in Figure 7, λ_c is a decreasing function of f_2 . It is clear that λ_c obviously decreases as f_2 ($f_2 \in (0, 1)$) increases and basically remains unchanged as f_2 ($f_2 \in (1, 2)$) increases. Figure 8 suggests that λ_c grows as β increases ($f_1 = 1.8, f_2 = 0.2$) and λ_c decreases as β increases ($f_1 = 0.2, f_2 = 1.8$), which means that the rapid spread of the epidemic can inhibit the outbreak of rumors and promote the dissemination of authoritative information. A smaller information forgetting rate, a greater rumor

FIGURE 5: The influence of f_1 on β_c .

forgetting rate, and a greater infectious disease recovery rate can raise the awareness threshold. Moreover, the curve of Model II is lower than that of Model I with equal parameters in both Figures 7 and 8, publicizing the epidemic information through the infected and cured people will contribute to spreading crisis awareness and preventing the outbreak of the epidemic.

4.2. Infection Scale. The key influencing factors for ρ_I and ρ_A are f_1 and f_2 , respectively. Figure 9 reveals that a larger epidemic cure rate μ helps control the epidemic diffusion scale and a smaller awareness forgetting rate σ contributes to the spread of awareness.

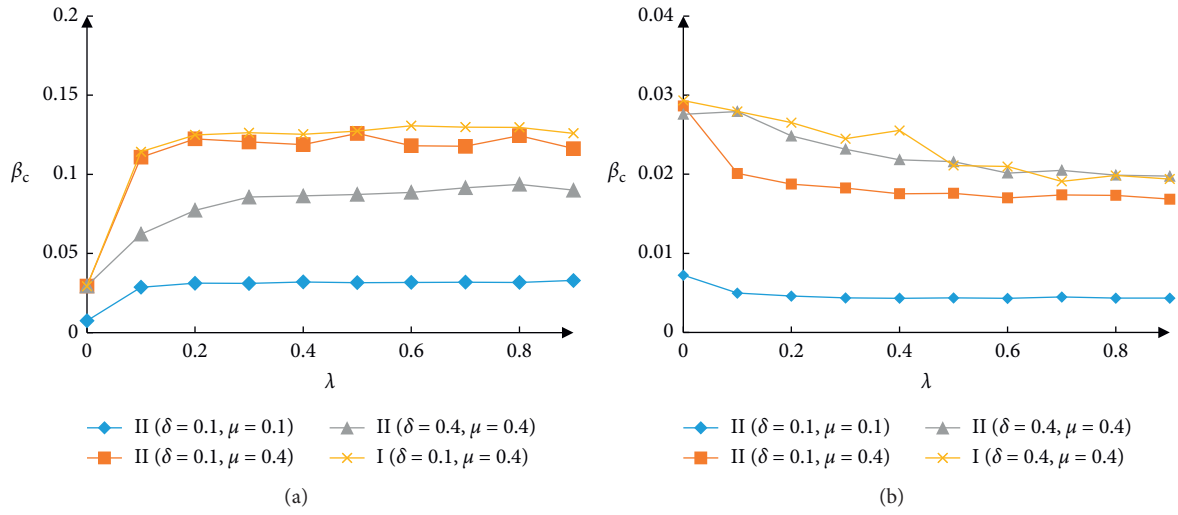


FIGURE 6: The influence of λ on β_c : (a) $f_1 = 0.2, f_2 = 1.8$; (b) $f_1 = 1.8, f_2 = 0.2$.

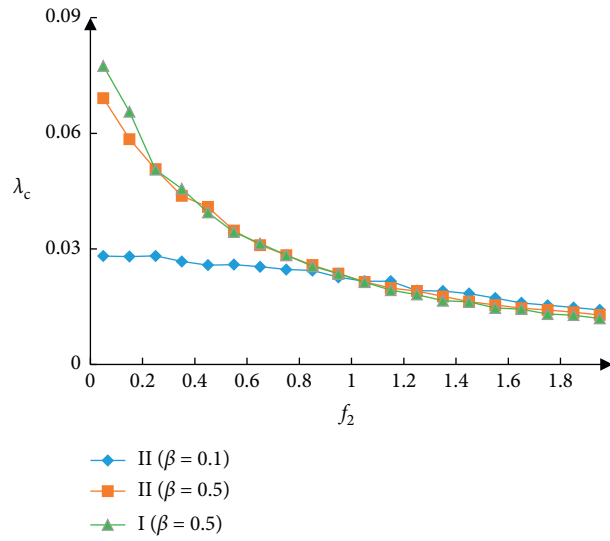


FIGURE 7: The influence of f_2 on λ_c .

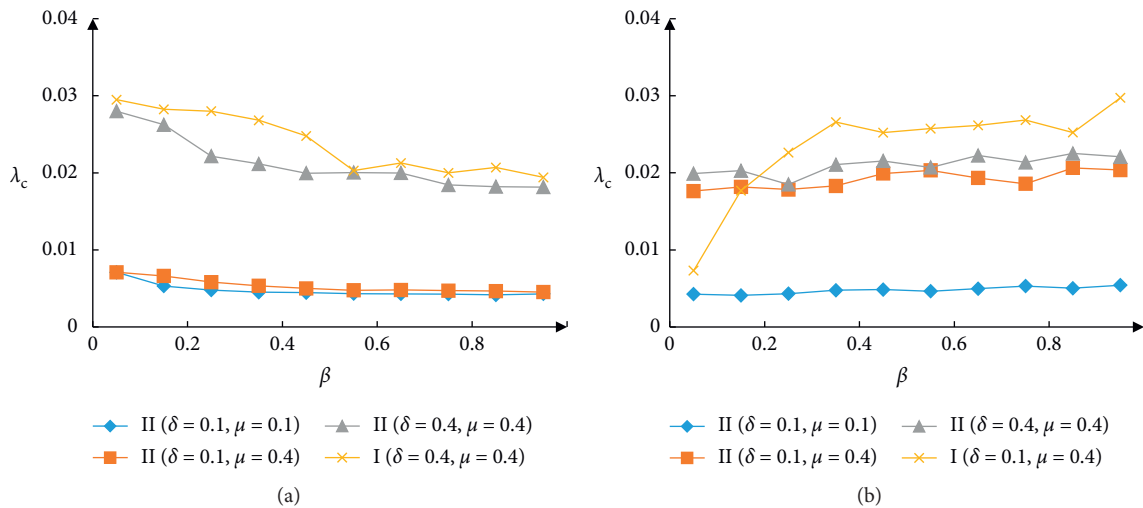


FIGURE 8: The influence of β on λ_c : (a) $f_1 = 0.2, f_2 = 1.8$; (b) $f_1 = 1.8, f_2 = 0.2$.

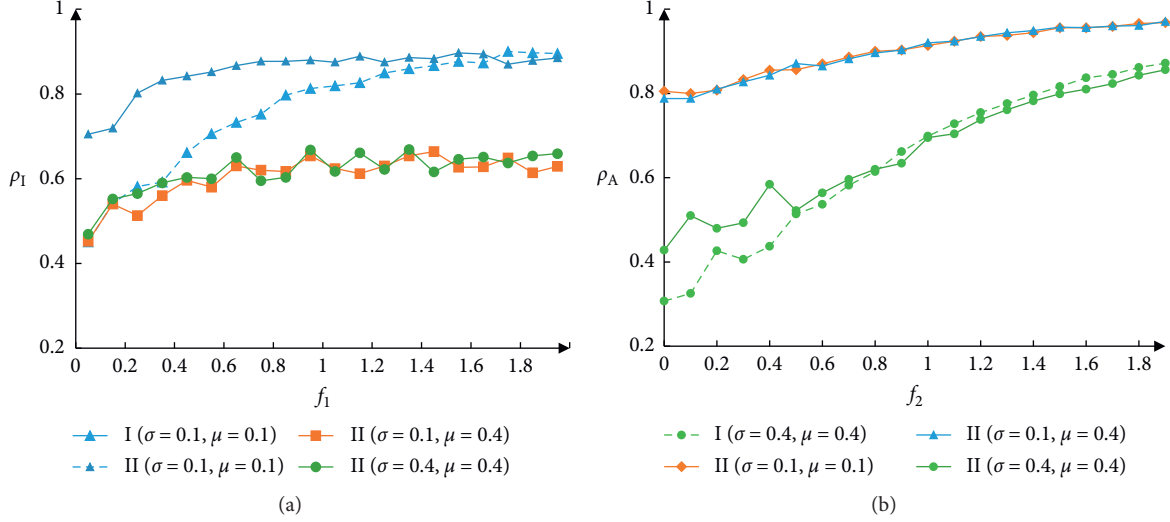


FIGURE 9: Impact of influence factors on the scale of infection.

4.2.1. Authoritative Information and Epidemic Interaction Model ($f_1 < 1, f_2 > 1$). Both ρ_A and ρ_I in Figure 10 show an upward trend with the increase of β , while the curve in Figure 10(a) is less obvious than that in Figure 10(b) because β can only indirectly affect the dissemination of authoritative information. In the meanwhile, a higher prevention degree of the individuals ($f_1 = 0.1$) corresponds to a smaller ρ_I (Figure 10(b)) and a smaller ρ_A (Figure 10(a)). Also, a stronger willingness of the individuals to spread information ($f_2 = 1.8$) led to a larger ρ_A and a smaller ρ_I .

The increase of λ contributes to the spread of awareness (Figure 11(a)) and the reduction of the epidemic scale (Figure 11(b)), which is in line with the actual situation. When the individuals' willingness to disseminate information gets stronger, ($f_2 = 1.8$), the value of ρ_A will be larger and ρ_I will decrease more. A higher prevention degree of the individuals ($f_1 = 0.1$) corresponds to a small ρ_I and a larger ρ_A .

The comparative experimental results of Figures 10 and 11 show that, under the same conditions, the ρ_A obtained by Model II is smaller and ρ_I is larger than that obtained by Model I, which suggests that models ignoring neighbors' behavior status would underestimate the epidemic scale and limit the formulation of epidemic control strategies.

4.2.2. Rumor and Epidemic Interaction Model ($f_1 > 1, f_2 < 1$). A larger β means a greater risk of individual infection, and it will lead to a larger ρ_I . However, rumors are not beneficial to epidemic control, and individuals would resist the spread of rumors; then, the curve of ρ_A would drop. As shown in Figure 12, a smaller f_2 ($f_2 = 0.1$) will lead to a smaller ρ_A , and a larger f_1 , ($f_1 = 1.8$) will cause a larger ρ_I .

A larger λ means that individuals are more likely to have rumor awareness, which will lead to a larger ρ_A . However, individuals in the epidemic layer affected by rumors will take improper behaviors to accelerate the spread of the epidemic, and the ρ_I curve will rise, as illustrated in Figure 13.

Furthermore, the more resistant an individual is to the rumor ($f_2 = 0.1$), the more difficult it is for the rumor to spread in the population, and ρ_A and ρ_I will be smaller; the greater an individual's contribution to epidemic transmission ($f_1 = 1.8$), the easier it is for the epidemic to spread, and ρ_A and ρ_I will be larger.

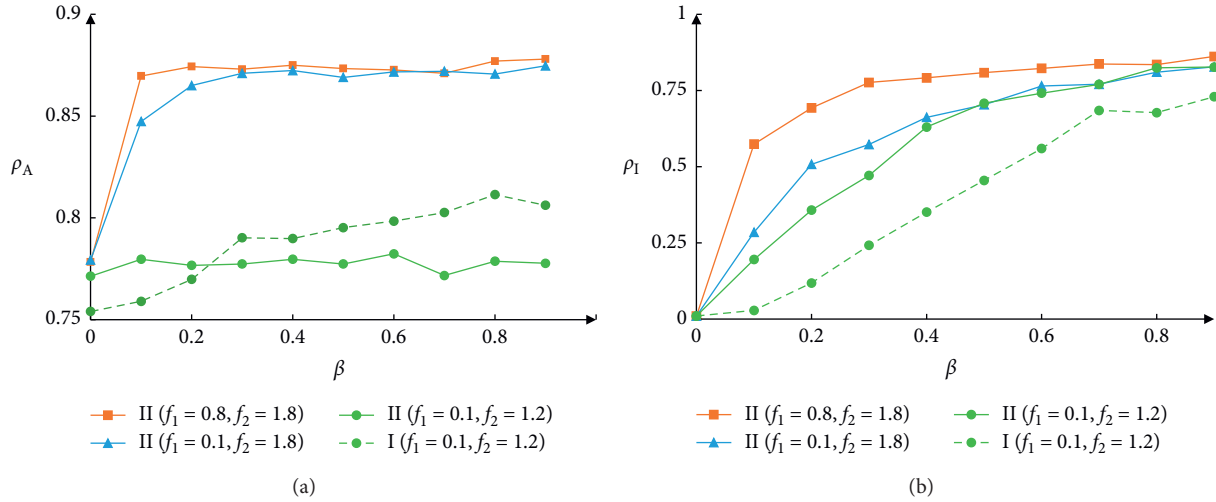
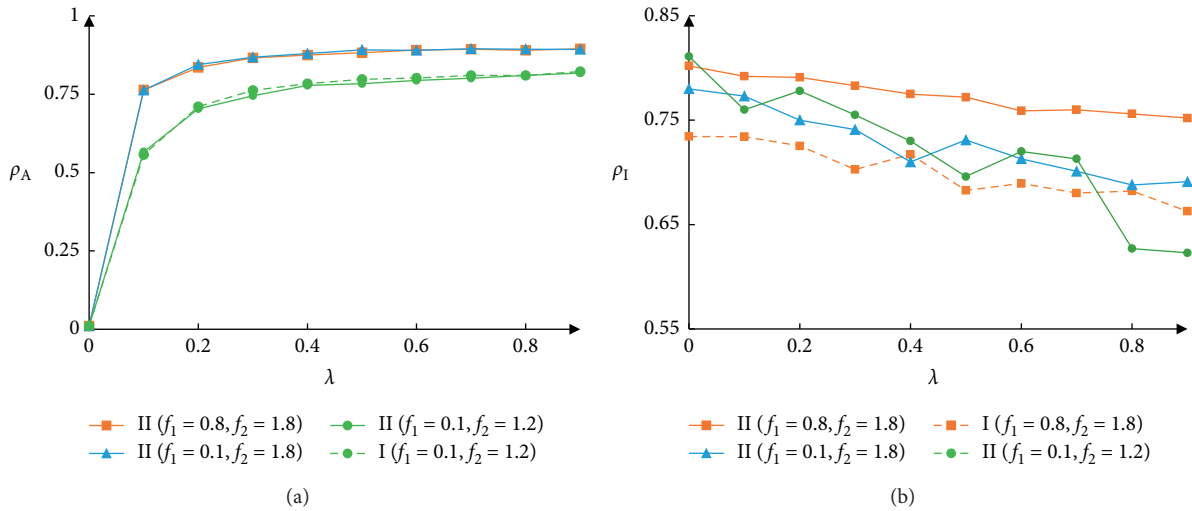
Figures 12 and 13 show that when the spread of the epidemic is accompanied by rumors, ρ_A obtained by Model II is larger than Model I, and the obtained ρ_I is smaller than Model I, which suggests that models ignoring neighbors' behavior status would overvalue the epidemic scale and influence the formulation of epidemic control strategies. As a result, it is necessary and reasonable to consider the influence of neighbors' behavior status when we establish an interactive propagation model.

5. Epidemic Analysis of COVID-19 in China

Four stages of strict measures for transmission containment were prompted during the early spread of COVID-19 in China, which reflects the change in public awareness. Each awareness stage plays discrepant roles in epidemic control and public response behaviors. We hypothesized that the awareness impact factor f_1 is negatively correlated with the importance of epidemic control measures and observe the assumed value of f_1 as shown in Table 3 [35].

Model I and Model II simulate the density of infected individuals as a function of t for different values of f_1 . It is clear that the large value of f_1 can increase the density of infected individuals, see Figure 14(a), and the neighbor behavior status can narrow the gap between the effects of different f_1 on the infection scale, see Figure 14(b). Therefore, the government can actively release positive information to enhance public awareness and restrict the behavior of infected neighbors for the purpose of controlling the spread of the epidemic during the COVID-19 epidemic.

Figure 15 is the daily new data released by the Publicity Department of the National Health Commission except for

FIGURE 10: The influence of β on ρ_A and ρ_I .FIGURE 11: The influence of λ on ρ_A and ρ_I .

Hubei Province, and Figure 16 is the data fitting curve of infected individuals (except Hubei Province). Figures 15 and 16 indicate that the number of infected individuals is gradually decreasing since January 29 with the strengthening of prevention and control measure [36].

Figure 17 shows the cumulative number of confirmed cases in Hubei Province from January to April. Since February 16, the cumulative number of confirmed cases has been close to 0, indicating that the epidemic has been under control. As we all know, public awareness is affected by the spread of the epidemic, and if the epidemic is effectively controlled, public awareness will become weaker. Public awareness is difficult to quantify, but we found such a phenomenon: the online big data information demonstrates that the amount of comprehensive information has risen rapidly since January 20, 2020, and reached the peak on February 15, and the peak time coincides with the epidemic

control time (Feb 16), as shown in Figure 17, which reflect the influence of epidemic spreading on the diffusion of awareness.

As shown in Table 4, the Hubei Provincial Government adjusts information report management at different stages of the epidemic. Each epidemic stage plays discrepant roles in public awareness diffusion. We hypothesized that the epidemic impact factor f_2 is positively correlated with the importance of information report management and observe the assumed value of f_2 .

Model I and Model II simulate the density of aware individuals as a function of t for different values of f_2 . It is clear that the large value of f_1 can increase the density of aware individuals, see Figure 18(a), and neighbor behavior status can expand the gap between the effects of different f_2 on the awareness scale, see Figure 18(b). Therefore, the measures taken by the government to publicize the epidemic

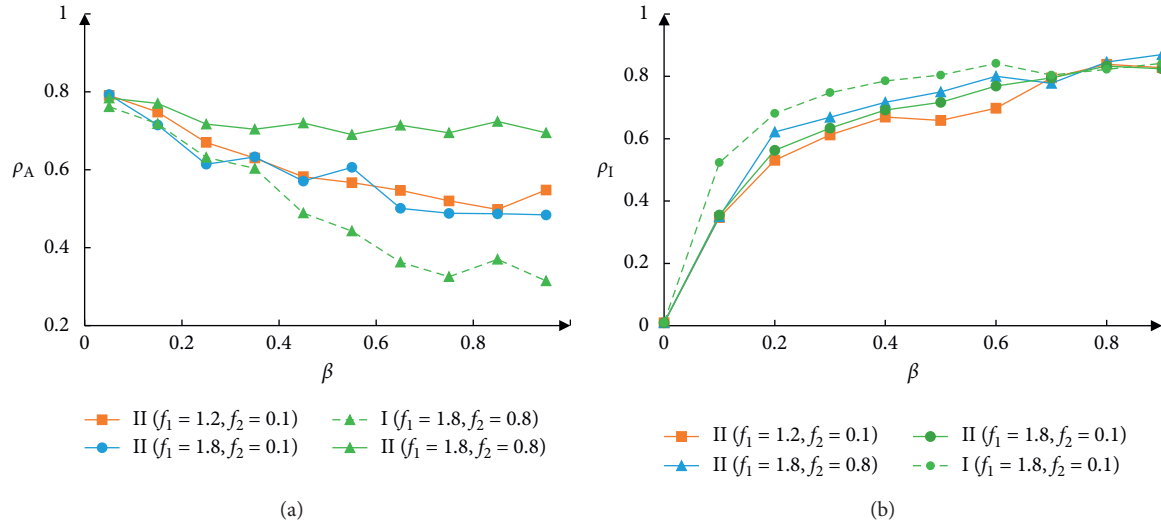


FIGURE 12: The influence of β on ρ_A and ρ_I .

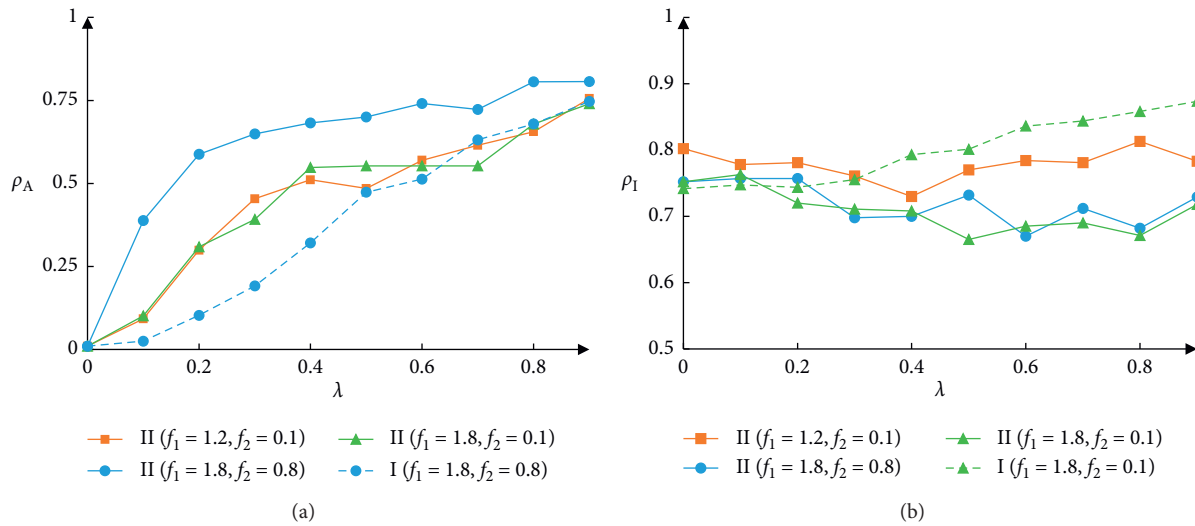


FIGURE 13: The influence of λ on ρ_A and ρ_I .

TABLE 3: Different government control measures and corresponding assumed value of f_1 .

| Phase | Date | Government measures | Importance | Assumed value of f_1 |
|-------|----------------------------------|---|-------------|------------------------|
| 1 | 29 December 2019–22 January 2020 | Early detection of the COVID-19 preliminary control | Significant | 0.8 |
| 2 | 23 January–29 January 2020 | (1) Public health level 1 response of 31 provinces (2) Strict exit screening (3) Medical support from other regions of China (4) Cancellation of mass gatherings (5) Methodological improvement on the diagnosis and treatment strategy | Critical | 0.6 |

TABLE 3: Continued.

| Phase | Date | Government measures | Importance | Assumed value of f_1 |
|-------|------------------------------|--|------------|------------------------|
| 3 | 30 January–11 February 2020 | (1) Public health level 1 response of 31 provinces (2) Strict exit screening (3) Domestic and international medical support (4) The larger scale of cancellation of mass gatherings (5) Further methodological improvement on the diagnosis and treatment strategy (6) Spontaneous household quarantine by citizens (7) Two newly built hospitals' put into use (8) A clinical trial of perspective medicines | Essential | 0.4 |
| 4 | 12 February–20 February 2020 | (1) Public health level 1 response of 31 provinces (2) Strict exit screening (3) Further medical support from home and abroad (4) Massive online teaching in a postponed semester (5) Orderly resumption of back to work (6) Addition of new diagnosis method—clinical diagnosis in Hubei Province (7) Interagency mechanism (8) Further exploration of an effective therapeutic strategy | Crucial | 0.2 |

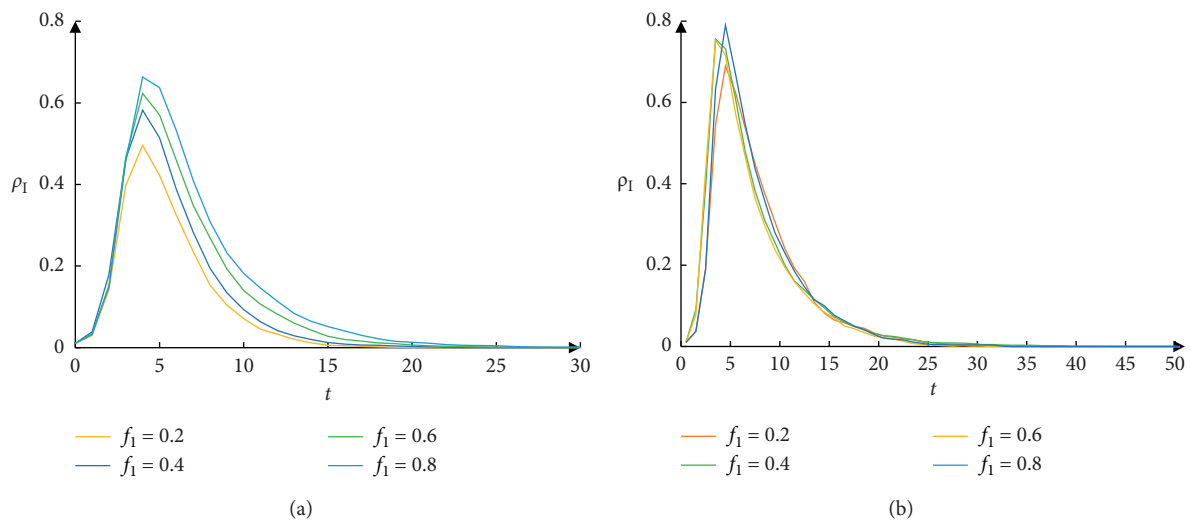


FIGURE 14: Changes in ρ_I under different levels of government measures: (a) Model I; (b) Model II.

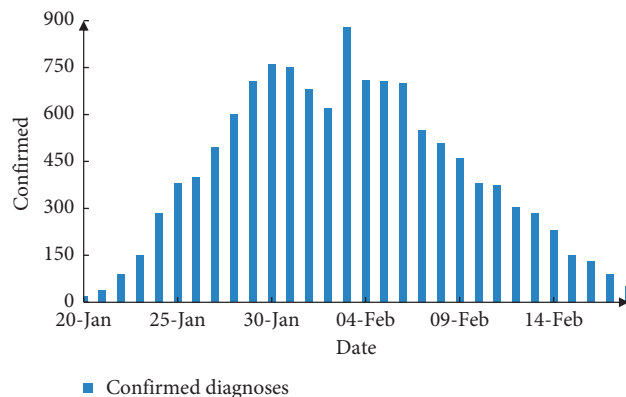


FIGURE 15: The number of newly confirmed diagnoses in regions outside Hubei.

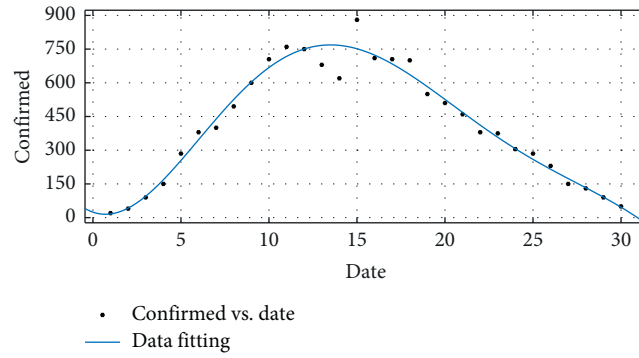


FIGURE 16: Date fitting by ploy regression.

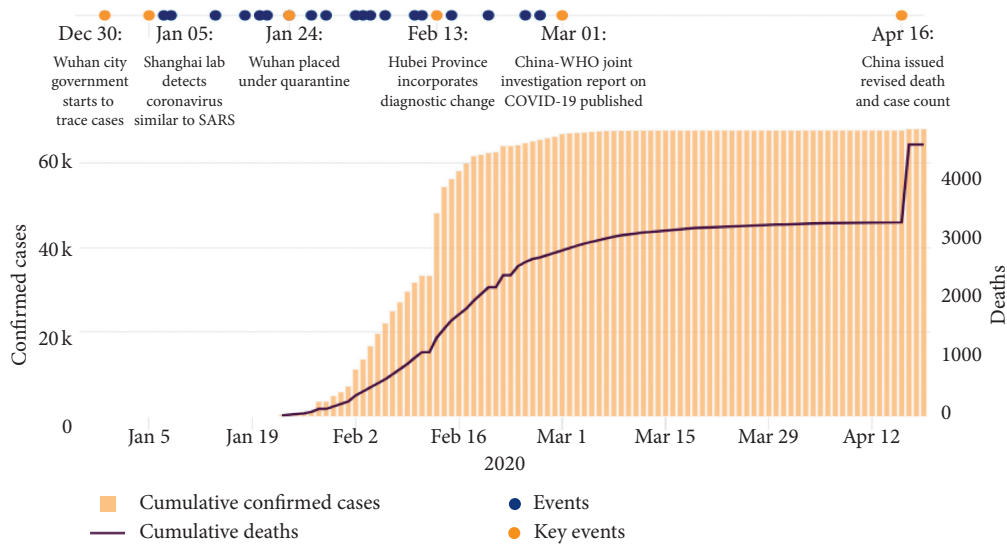


FIGURE 17: Accumulated timetable for confirmed cases in Hubei Province (<https://coronavirus.jhu.edu/map.html>).

TABLE 4: Information report of COVID-19 in Hubei Province and the corresponding assumed value of f_2 .

| Phase | Epidemic stage | Information report management changes | Importance | Assumed value of f_2 |
|-------|------------------|---|-------------|------------------------|
| 1 | Early period | (1) Holding special meetings (2) Establishment of emergency response teams (3) Issue guidance notice | Significant | 1.2 |
| 2 | Beginning period | (1) CDC has issued guidance documents for many times (2) Video training (3) Answer questions by telephone (4) Issuing guidance notices | Critical | 1.5 |
| 3 | High-risk period | (1) Provide 4 analysis reports of more than 20 pages per day (2) Analyze the characteristics of the disease (3) To assess the trend of the epidemic (4) Provide data support for leaders' decision-making | Crucial | 1.8 |
| 4 | Low-risk period | (1) Timely update of the COVID-19 monitoring analysis report (2) Pay attention to the detailed source and detection route of case information (3) Focus on asymptomatic infected persons and epidemic situation analysis at home and abroad | Essential | 1.6 |

*The information of COVID-19 in the study was mainly obtained from the National Health Commission of the People's Republic of China, Chinese Center for Disease Control and Prevention, WHO, Hubei Provincial for Disease Control and Prevention, and various websites of Chinese government agencies and official media, as well as some previous studies.

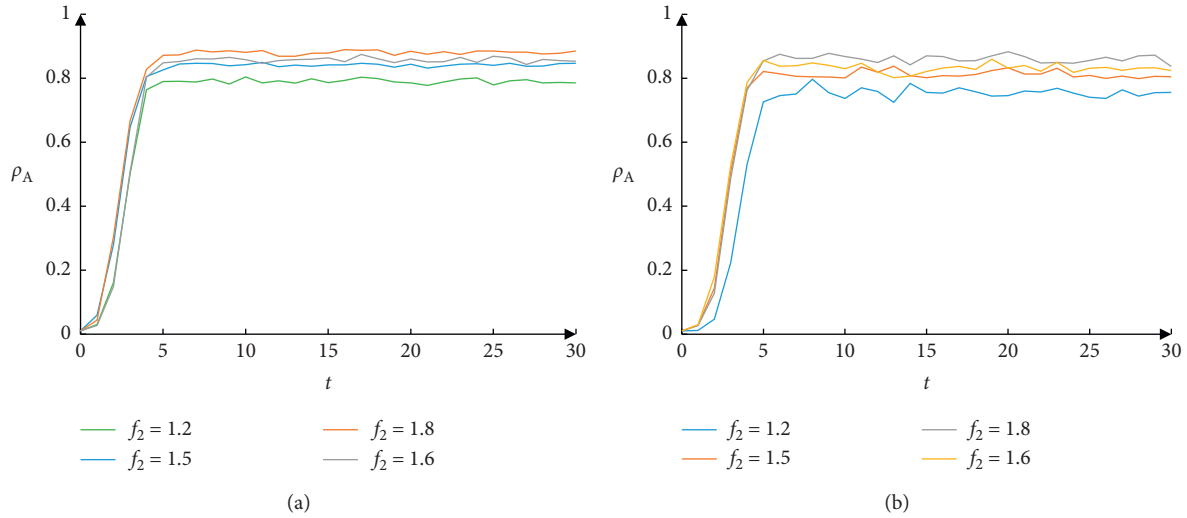


FIGURE 18: Changes in ρ_A under different epidemic stages: (a) Model I; (b) Model II.

data and epidemic prevention process during the spread of COVID-19 can enhance public awareness and effectively control the spread of the epidemic.

6. Conclusions

In this paper, on the one hand, we discuss the interaction between awareness diffusion (including authoritative information and rumors) and epidemic spreading; on the other hand, we analyze the influence of neighbor status on epidemic spreading when awareness diffusion and neighbor behavior are coupled in multiple networks. Distinct from previous studies, people can get epidemic prevention information to reduce the infection rates from their social communication circles, such as microblogs, as well as increase awareness based on their own health status. Moreover, different neighbor status play distinct roles in the awareness acquisition process and the response behavior change process of an individual. We investigate the impact of these factors on epidemic spreading processes and obtain the epidemic threshold with the MMCA approach. Analysis based on the numerical simulations reveal that the epidemic can be reduced by publishing authoritative information to reduce f_1 , publicizing the epidemic prevention process to increase f_2 , and controlling neighbor behavior to be more reasonable, for example, isolating infected neighbors and encouraging them to propagate information related to the epidemic.

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 no conflicts of interest.

Acknowledgments

This work was supported by the Education Project of Zhejiang Province (No. Y202045064).

References

- [1] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *Journal of Autoimmunity*, vol. 109, Article ID 102433, 2020.
- [2] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, "COVID-19 infection: origin, transmission, and characteristics of human coronaviruses," *Journal of Advanced Research*, vol. 24, pp. 91–98, 2020.
- [3] C.-J. Fan, Y. Jin, L.-A. Huo, C. Liu, Y.-P. Yang, and Y.-Q. Wang, "Effect of individual behavior on the interplay between awareness and disease spreading in multiplex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 461, pp. 523–530, 2016.
- [4] X.-X. Zhan, C. Liu, G. Zhou, Z.-K. Zhang et al., "Coupling dynamics of epidemic spreading and information diffusion on complex networks," *Applied Mathematics and Computation*, vol. 332, pp. 437–448, 2018.
- [5] K. M. A. Kabir, K. Kuga, and J. Tanimoto, "Analysis of SIR epidemic model with information spreading of awareness," *Chaos, Solitons & Fractals*, vol. 119, pp. 118–125, 2019.
- [6] Q. Guo, X. Jiang, and Y. Lei, "Two-stage effects of awareness cascade on epidemic spreading in multiplex networks," *Physical Review E*, vol. 91, no. 1, Article ID 012822, 2015.
- [7] S. Pei, L. Muchnik, and J. S. Andrade Jr., "Searching for super spreaders of information in real-world social media," *Scientific Reports*, vol. 4, p. 5547, 2014.
- [8] S. Funk, E. Gilad, and V. A. A. Jansen, "Endemic disease, awareness, and local behavioural response," *Journal of Theoretical Biology*, vol. 264, no. 2, pp. 501–509, 2010.
- [9] Z. Wang, M. A. Andrews, Z.-X. Wu, L. Wang, and C. T. Bauch, "Coupled disease-behavior dynamics on complex networks: a review," *Physics of Life Reviews*, vol. 15, pp. 1–29, 2015.
- [10] K. M. A. Kabir, K. Kuga, and J. Tanimoto, "Effect of information spreading to suppress the disease contagion on the

- epidemic vaccination game,” *Chaos, Solitons & Fractals*, vol. 119, pp. 180–187, 2019.
- [11] K. A. Kabir, K. Kuga, and J. Tanimoto, “The impact of information spreading on epidemic vaccination game dynamics in a heterogeneous complex network—a theoretical approach,” *Chaos, Solitons & Fractals*, vol. 132, Article ID 109548, 2020.
- [12] C. Granell, S. Gomez, and A. Arenas, “Dynamical interplay between awareness and epidemic spreading in multiplex networks,” *Physical Review Letters*, vol. 111, no. 12, Article ID 128701, 2013.
- [13] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen, “The spread of awareness and its impact on epidemic outbreaks,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6872–6877, 2009.
- [14] W. Wang, Q. H. Liu, and S. M. Cai, “Suppressing disease spreading by using information diffusion on multiplex networks,” *Scientific Reports*, vol. 6, no. 1, pp. 1–14, 2016.
- [15] W. Wang, Q.-H. Liu, J. Liang, Y. Hu, and T. Zhou, “Co-evolution spreading in complex networks,” *Physics Reports*, vol. 820, pp. 1–51, 2019.
- [16] P. Zhu, X. Wang, Q. Zhi, J. Ma, and Y. Guo, “Analysis of epidemic spreading process in multi-communities,” *Chaos, Solitons & Fractals*, vol. 109, pp. 231–237, 2018.
- [17] P. Zhu, X. Wang, S. Li, Y. Guo, and Z. Wang, “Investigation of epidemic spreading process on multiplex networks by incorporating fatal properties,” *Applied Mathematics and Computation*, vol. 359, pp. 512–524, 2019.
- [18] K. M. A. Kabir and J. Tanimoto, “Analysis of epidemic outbreaks in two-layer networks with different structures for information spreading and disease diffusion,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 72, pp. 565–574, 2019.
- [19] L. Y. Shang, “Discrete-time epidemic dynamics with awareness in random networks,” *International Journal of Biomathematics*, vol. 6, no. 2, pp. 147–153, 2013.
- [20] Y. Shang, “Modeling epidemic spread with awareness and heterogeneous transmission rates in networks,” *Journal of Biological Physics*, vol. 39, no. 3, pp. 489–500, 2013.
- [21] F. Bagnoli, P. Lio, and L. Sgauri, “Risk perception in epidemic modeling,” *Physical Review E*, vol. 76, no. 6, Article ID 061904, 2007.
- [22] J. Alstott, P. Panzarasa, and M. Rubinov, “A unifying framework for measuring weighted rich clubs,” *Scientific Reports*, vol. 4, no. 1, pp. 1–6, 2014.
- [23] Q. Wu, X. Fu, and M. Small, “The impact of awareness on epidemic spreading in networks,” *Chaos: an interdisciplinary journal of nonlinear science*, vol. 22, no. 1, Article ID 013101, 2012.
- [24] Z. Wang, C. T. Bauch, S. Bhattacharyya et al., “Statistical physics of vaccination,” *Physics Reports*, vol. 664, pp. 1–113, 2016.
- [25] Z. Li, P. Zhu, and D. Zhao, “Suppression of epidemic spreading process on multiplex networks via active immunization,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 7, Article ID 073111, 2019.
- [26] J.-Q. Kan and H.-F. Zhang, “Effects of awareness diffusion and self-initiated awareness behavior on epidemic spreading—an approach based on multiplex networks,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 44, pp. 193–203, 2017.
- [27] Q. Guo, Y. Lei, and C. Xia, “The role of node heterogeneity in the coupled spreading of epidemics and awareness,” *PLoS One*, vol. 11, no. 8, Article ID e0161037, 2016.
- [28] P. Zhu, Q. Zhi, Y. Guo, and Z. Wang, “Analysis of epidemic spreading process in adaptive networks,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 7, pp. 1252–1256, 2018.
- [29] X. Chen, K. Gong, R. Wang, S. Cai, and W. Wang, “Effects of heterogeneous self-protection awareness on resource-epidemic coevolution dynamics,” *Applied Mathematics and Computation*, vol. 385, Article ID 125428, 2020.
- [30] Y. Pan and Z. Yan, “The impact of individual heterogeneity on the coupled awareness-epidemic dynamics in multiplex networks,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 6, Article ID 063123, 2018.
- [31] M. C. M. Balemans, M. M. H. Huibers, N. W. D. Eikelenboom et al., “Reduced exploration, increased anxiety, and altered social behavior: autistic-like features of euchromatin histone methyltransferase 1 heterozygous knockout mice,” *Behavioural Brain Research*, vol. 208, no. 1, pp. 47–55, 2010.
- [32] H. Faneca, V. A. Figueiredo, I. Tomaz et al., “Vanadium compounds as therapeutic agents: some chemical and biochemical studies,” *Journal of Inorganic Biochemistry*, vol. 103, no. 4, pp. 601–608, 2009.
- [33] K. Hukushima and K. Nemoto, “Exchange Monte Carlo method and application to spin glass simulations,” *Journal of the Physical Society of Japan*, vol. 65, no. 6, pp. 1604–1608, 1996.
- [34] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, “New approach to Monte Carlo calculation of the free energy: method of expanded ensembles,” *The Journal of Chemical Physics*, vol. 96, no. 3, pp. 1776–1783, 1992.
- [35] Y. Fang, Y. Nie, and M. Penny, “Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis,” *Journal of Medical Virology*, vol. 92, no. 6, pp. 645–659, 2020.
- [36] Y. C. Chen, P. E. Lu, and C. S. Chang, “A time-dependent SIR model for COVID-19 with undetectable infected persons,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 3279–3294, 2020.

Research Article

Spreading Dynamics of a 2SIH2R, Rumor Spreading Model in the Homogeneous Network

Yan Wang,^{1,2} Feng Qing ,^{1,2} Jian-Ping Chai,^{1,2} and Ye-Peng Ni^{1,2}

¹State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China

²School of Data Science and Intelligent Media, Communication University of China, Beijing 100024, China

Correspondence should be addressed to Feng Qing; qing329364213@cuc.edu.cn

Received 15 October 2020; Revised 5 January 2021; Accepted 29 January 2021; Published 10 February 2021

Academic Editor: xiaoke xu

Copyright © 2021 Yan Wang 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.

With the rapid development of social network in recent years, the threshold of information dissemination has become lower. Most of the time, rumors, as a special kind of information, are harmful to society. And once the rumor appears, the truth will follow. Considering that the rumor and truth compete with each other like light and darkness in reality, in this paper, we study a rumor spreading model in the homogeneous network called 2SIH2R, in which there are both spreader1 (people who spread the rumor) and spreader2 (people who spread the truth). In this model, we introduced discernible mechanism and confrontation mechanism to quantify the level of people's cognitive abilities and the competition between the rumor and truth. By mean-field equations, steady-state analysis, and numerical simulations in a generated network which is closed and homogeneous, some significant results can be given: the higher the discernible rate of the rumor, the smaller the influence of the rumor; the stronger the confrontation degree of the rumor, the smaller the influence of the rumor; the larger the average degree of the network, the greater the influence of the rumor but the shorter the duration. The model and simulation results provide a quantitative reference for revealing and controlling the spread of the rumor.

1. Introduction

With the continuous emergence of social media platforms, the traditional media era has gradually turned into the self-media era, and information dissemination has become faster, wider in scope, and deeper than ever [1]. Rumors, as a special kind of information, have greatly increased the possibility of artificial release of rumors due to their own confusion, timeliness, and psychological satisfaction to the people who spread the rumor. Coupled with the self-media era, the threshold for spreading rumors is further lowered [2]. In today's society, there are some people who use people's public psychology to create rumors to obtain benefits from it [3, 4]. This behavior will cause public panic and harm society. Therefore, in order to reveal the law of rumors dissemination and reduce the negative impact of rumors on society, it is necessary to establish a suitable mathematical model to analyse the characteristics and mechanisms of rumors dissemination process.

In the 1960s, Daley and Kendall [5] proposed the DK model which is a classic model in the field of rumor spreading. The model divides the population into three categories: people who have never heard of rumors (ignorant), people who spread rumors (spreader), and people who have heard the rumors but do not spread (stifler). In order to characterize the reaction after receiving rumors, individuals will transform their identities in the three categories based on probability. Zanette [6, 7] used complex network theories to study how the rumors spread. He established a rumor spreading model in small-world networks and proved the threshold of rumor spreading. Moreno [4] established a SIR (susceptible-infective-refractory) model in a scale-free network and analysed the simulation results. In the process of research, some scholars adjusted the SIR model according to the research purpose and applied it to the complex network getting numerous great consequences. Wan [8] studied the propagation process of the adjusted SIR model in a homogeneous network and proposed two strategies for

network rumor immunity: active immunity and passive immunity. Zhao [9] introduced a media report mechanism and studied the influence of media reports on rumors. Askarizadeh [10] introduced an anti-rumor mechanism and proposed a game model to analyse the process of rumor spread in social networks and concluded that anti-rumor will affect the spread of rumors. Zhang [11] established I2S2R dynamic rumor propagation models in homogeneous and heterogeneous networks. Huo [12] proposed that the SIBInIu model divides the population into four categories and concluded that the losing-interest rate and stifling rate have a negative impact on the scale of the final spread of rumors. Deng [13] introduced the forgetting and memory mechanisms in the process of studying the spread of rumors. Gu [14] established a SEIR model on Facebook's user data set and concluded that acquaintance immunization is the best solution to curb online rumors through the comparison of multiple immunization strategies. Wang [15] established the SIRaRu model and proposed that when the ignorant meets the spreader, the ignorant believes the rumor or not with probability, corresponding to the formation of two stifle states. In 2014, Wang [16] also considered the existence of multiple rumors in a network, and one type of rumors would be affected by another type of rumors. Wang [17], Yang [18], and Xia [19] introduced the hesitation mechanism in their respective models. People who heard the rumors temporarily did not spread them; when they heard the rumors again, they became the spreaders with probability. Based on the DK model, Huo [20] divided spreaders into two categories, namely, spreaders with high activity and spreaders with low activity. Later, Huo [21] introduced the indiscernible degree mechanism in the model to describe the individual's cognitive ability. Ran [22] introduced a rumor rejection mechanism while considering the impact of individual differences on the spread of rumors and established an IWSR rumor spread model. Dong [23] proposed a double-identity rumor spreading model; that is, in addition to ignorant, spreader, and stifler, each network node also has one of three other identities, namely, rumor creator, rumor controller, and normal user. Zan [24] established the SICR model to introduce the counterattack mechanism of rumors; that is, when the spreader contacts the counterattack, the spreader becomes a stifler with probability.

In the above studies, many have made great contributions to the theoretical research on the process of rumor spreading on complex networks. However, there are two shortcomings in the theory that need to be improved. The first one is that, in reality, the discernible degree of the rumor is an important variable, but most previous studies did not quantify this. Allport and Postman [25] believe that there are three conditions for the generation and spread of rumors: the first one is the lack of information; the second one is people's anxiety; the third one is that the society is in crisis. Based on this, they proposed a classical formula: rumors = $i \times a$ (where i represents the importance of information and a represents the degree of unknowability of the event). The other improvement to be made is that no research has been done on the spread of truth, the opposite of rumors. With the rumor, there is also the truth. In reality,

there are always some wise men who can reveal the rumor and spread the truth, in which time there will be a confrontation relationship between the rumors and the truth [24]. Based on this, we divide the population into six categories, people who have never heard of rumors or truth (ignorant), people who spread rumors (spreader1), people who spread truth (spreader2), people who have heard the rumors but do not spread temporarily (hesitant1), people who have heard the rumors but do not spread (stifler1), people who have heard the truth but do not spread (stifler2), and propose the 2SIH2R model with the discernible mechanism and the confrontation mechanism.

The organization of the paper is the following. In Section 2, the 2SIH2R model is defined, and the mean-field equations of the model are established in the homogeneous network [16, 26]. In Section 3, we study the rumor spreading threshold of model propagation by changing initial conditions and parameters and extend the spreading threshold under special circumstances [27] to general conditions. In Section 4, through simulation, we study the influence of discernible mechanism, confrontation mechanism, and average degree on the rumor. In Section 5, conclusions of the paper and future work are given.

2. 2SIH2R Rumor Spreading Model

We consider a closed and mixed population composed of N individuals as a complex network, where individuals and their contacts can be represented by vertexes and edges. This network can be described by an undirected graph $G = (V, E)$, where V denotes the vertexes and E represents the edges. At each time t , the people in the network can be divided into S_1, S_2, I, H, R_1, R_2 , separately, representing people who spread the rumor, people who spread truth, people who have never heard of the rumor or truth, people who have heard the rumor but do not spread temporarily, people who have heard the rumor but do not spread, and people who have heard the truth but do not spread. The rumor spreading process of the 2SIH2R can be seen in Figure 1.

In Figure 1, the solid/dotted line from the "Ignorant" represents that the ignorant contacts with spreader1/spreader2, and the rumor spreading rules of the 2SIH2R model can be summarized as follows.

- (1) We use m to describe the discernible rate of the rumor, and $f(m)$ to describe people's ability to reveal rumors. The function of f is to map the characteristics of the rumor to the characteristics of the people. The greater the m , the greater the probability that the rumor will be revealed. The greater the $f(m)$, the greater the probability that the people will not believe the rumor immediately. We assume that there is a positive correlation between m and $f(m)$.
- (2) When an ignorant person meets a spreader1: (a) the ignorant may believe the rumor and spread it with probability $(1 - f(m))\lambda_1$. The λ_1 is rumor spreading rate; (b) the ignorant may not believe the rumor

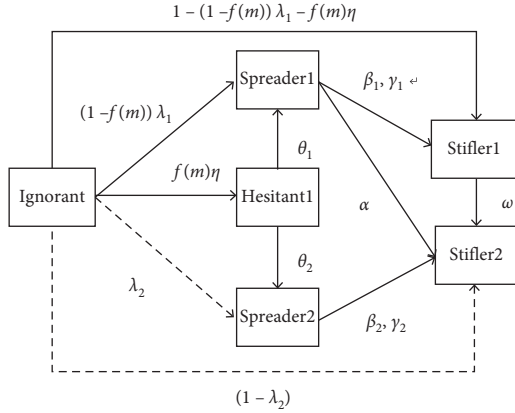


FIGURE 1: 2SIH2R rumor spreading process.

immediately and hesitate to spread it with probability $f(m)\eta$. The η is potential spreading rate; (c) the ignorant may have no interest of the rumor with probability $1 - (1 - f(m))\lambda_1 - f(m)\eta$.

- (3) When an ignorant person meets a spreader2: (a) the ignorant may believe the truth and spread it with probability λ_2 , called truth spreading rate; (b) the ignorant may have no interest of the truth with probability $(1 - \lambda_2)$. Since the truth is generally issued by an authoritative organization, or there is evidence to support it, the truth is relatively more objective, accurate, and clear. It is easier for the ignorant to judge, and not easy to become a hesitant.
- (4) We consider that the hesitant1 have a desire to spread the information, because of the suspicion of the rumor and the environmental impact when they receive the rumor; they do not spread the rumor immediately. In the time of hesitation, hesitant1 may believe the rumor to spread it, or they may discover the truth and spread it. So, we assume that, at each step, the hesitant1 will spontaneously become people who spread the rumor (spreader1) with probability θ_1 , and people who spread truth (spreader2) with probability θ_2 .
- (5) When a spreader1 (spreader2) encounters another spreader1 (spreader2), he/she could think the rumor (truth) is widely known. So, the spreader1 (spreader2) may lose spreading enthusiasm and become a stifler1 (stifler2) with rumor (truth) losing-interest rate β_1 (β_2).
- (6) At each step, spreader1 (spreader2) becomes a stifler1 (stifler2) spontaneously with probability γ_1 (γ_2), called rumor (truth) forgetting-rate.
- (7) At each step, the stifler1 will spontaneously become stifler2 with probability ω , because of the improvement of their own cognitive level.
- (8) When a spreader1 encounters spreader2, the spreader1 will believe the truth rather than the rumor with probability α , because of the confrontation mechanism between the truth and rumor. The α is confrontation rate.

Moreover, the 2SIH2R model is applied to a generated network which is a closed and homogeneous population consisting of N individuals [17, 28]. We use $S_1(t)$, $S_2(t)$, $I(t)$, $H(t)$, $R_1(t)$, $R_2(t)$, separately, to represent the densities of spreader1, spreader2, ignorant, hesitant1, stifler1, and stifler2, and at any step, the following condition always exists:

$$I(t) + S_1(t) + S_2(t) + H(t) + R_1(t) + R_2(t) = 1. \quad (1)$$

According to the rumor spreading rules, the mean-field equation of 2SIH2R model can be expressed as follows:

$$\frac{dI(t)}{dt} = -\langle k \rangle (S_1(t) + S_2(t))I(t), \quad (2)$$

$$\begin{aligned} \frac{dS_1(t)}{dt} &= (1 - f(m))\lambda_1 \langle k \rangle S_1(t)I(t) + \theta_1 H(t) \\ &\quad - \alpha \langle k \rangle S_1(t)S_2(t) \\ &\quad - \beta_1 \langle k \rangle S_1(t)(S_1(t) + R_1(t) + H(t)) - \gamma_1 S_1(t), \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{dS_2(t)}{dt} &= \lambda_2 \langle k \rangle S_2(t)I(t) + \theta_2 H(t) - \beta_2 \langle k \rangle S_2(t)(S_2(t) \\ &\quad + R_2(t)) - \gamma_2 S_2(t), \end{aligned} \quad (4)$$

$$\frac{dH(t)}{dt} = f(m)\eta \langle k \rangle S_1(t)I(t) - (\theta_1 + \theta_2)H(t), \quad (5)$$

$$\begin{aligned} \frac{dR_1(t)}{dt} &= (1 - (1 - f(m))\lambda_1 - f(m)\eta) \langle k \rangle S_1(t)I(t) \\ &\quad + \beta_1 \langle k \rangle S_1(t)(S_1(t) + R_1(t) + H(t)) + \gamma_1 S_1(t) \\ &\quad - \omega R_1(t), \end{aligned} \quad (6)$$

$$\begin{aligned} \frac{dR_2(t)}{dt} &= (1 - \lambda_2) \langle k \rangle S_2(t)I(t) + \alpha \langle k \rangle S_1(t)S_2(t) \\ &\quad + \beta_2 \langle k \rangle S_2(t)(S_2(t) + R_2(t)) + \gamma_2 S_2(t) + \omega R_1(t), \end{aligned} \quad (7)$$

where $\langle k \rangle$ represents the average degree of the generated network.

3. Steady-State Analysis

In this section, we will consider the three situations of the model. When the system reaches the steady state, there is neither spreader1 nor spreader2. So, we can give the condition in the final state: $S_1 = \lim_{t \rightarrow \infty} S_1(t) = 0$, $S_2 = \lim_{t \rightarrow \infty} S_2(t) = 0$, $H = \lim_{t \rightarrow \infty} H(t) = 0$, and $\lim_{t \rightarrow \infty} (I(t) + R_1(t) + R_2(t)) = 1$. It is assumed that $I = \lim_{t \rightarrow \infty} I(t)$, $R_1 = \lim_{t \rightarrow \infty} R_1(t)$, $R_2 = \lim_{t \rightarrow \infty} R_2(t)$. The final size of the rumor (truth) R_1 (R_2) will be calculated to measure the level of the rumor (truth) influence [12], and $R = R_1 + R_2$ is used to measure the level of influence of the model. We will study the rumor spreading threshold of the model by analysing the final size of R . The sum of equations (6) and (7) is divided by equation (2); we have

$$\frac{dR(t)}{dI(t)} = \frac{d(R_1(t) + R_2(t))}{dI(t)} = \frac{(1 - (1 - f(m))\lambda_1 - f(m)\eta)S_1(t) + (1 - \lambda_2)S_2(t)}{S_1(t) + S_2(t)} - \frac{\beta_1 S_1(t)(S_1(t) + R_1(t) + H(t)) + \beta_2 S_2(t)(S_2(t) + R_2(t)) + \alpha S_1(t)S_2(t)}{(S_1(t) + S_2(t))I(t)} - \frac{\gamma_1 S_1(t) + \gamma_2 S_2(t)}{\langle k \rangle (S_1(t) + S_2(t))I(t)} \quad (8)$$

3.1. Steady-State Analysis of Rumor. At the beginning of model spreading, in this situation, we assumed that there is only one spreader1 who spreads the rumor, and there is no truth. So, the initial condition can be given: $S_2(0) = 0$, $\omega = 0$, $S_1(0) = (1/N) \approx 0$, $I(0) = (N - 1/N) \approx 1$, $H(0) = R_1(0) = R_2(0) = 0$. After a while, the number of

spreader1 will increase to the top; then, it reduces to zero, at which time the system reaches stability.

Since $S_2(0) = 0$, $\omega = 0$, equations (4) and (7), we can know $S_2(t) = 0$ and $R_2(t) = 0$. So, there is the normalization condition $I(t) + S_1(t) + H(t) + R_1(t) = 1$. Considering the above-mentioned conditions, equation (8) becomes

$$\begin{aligned} \frac{dR(t)}{dI(t)} &= \frac{d(R_1(t) + R_2(t))}{dI(t)} = -(1 - (1 - f(m))\lambda_1 - f(m)\eta) - \frac{\beta_1(1 - I(t))}{I(t)} - \frac{\gamma_1}{\langle k \rangle I(t)} \\ &= -c - \frac{\beta_1(1 - I(t))}{I(t)} - \frac{\gamma_1}{\langle k \rangle I(t)} \Rightarrow dR(t) = -cdI(t) - \frac{\beta_1(1 - I(t))}{I(t)}dI(t) - \frac{\gamma_1}{\langle k \rangle I(t)}dI(t) \\ \Rightarrow \int_0^\infty dR(t)dt &= \int_0^\infty \left(-cdI(t) - \frac{\beta_1(1 - I(t))}{I(t)}dI(t) - \frac{\gamma_1}{\langle k \rangle I(t)}dI(t) \right) dt \\ \Rightarrow R &= (\beta_1 - c)(I - 1) - \left(\beta_1 + \frac{\gamma_1}{\langle k \rangle} \right) \ln I \Rightarrow (\beta_1 - c + 1)R = - \left(\beta_1 + \frac{\gamma_1}{\langle k \rangle} \right) \ln(1 - R) \\ \Rightarrow \frac{\beta_1 - c + 1}{- \left(\beta_1 + \frac{\gamma_1}{\langle k \rangle} \right)} R &= \ln(1 - R) \Rightarrow R = 1 - e^{- \left(\beta_1 - c + 1 / \beta_1 + (\gamma_1 / \langle k \rangle) \right) R} \Rightarrow R = 1 - e^{-\varepsilon R}, \end{aligned} \quad (9)$$

where $c = 1 - (1 - f(m))\lambda_1 - f(m)\eta$ and $\varepsilon = (\beta_1 - c + 1/\beta_1 + (\gamma_1/\langle k \rangle))$. Only when $\varepsilon > 1$ will equation (9) have a non-zero solution. For $f(m) \neq 1$, we have

$$\lambda_{1c} = \frac{\gamma_1 - \langle k \rangle f(m)\eta}{\langle k \rangle (1 - f(m))}. \quad (10)$$

Due to the confrontation mechanism, when $0 \leq \lambda_1 \leq \lambda_{1c}$, the rumor must not spread widely in the generated network.

3.2. Steady-State Analysis of Truth. It is assumed that the government or authoritative media have already begun to spread the truth before a rumor event occurs. Then, when a rumor event occurs, there will be no rumor spreader in the population. So, in this situation, the initial condition can be given: $S_1(0) = 0$. Then, we can prove $S_1(t) = 0$ from equation (4), and $R_1(t) = 0$ from equation (6). Since there are no hesitant in the network at the beginning, we can prove $H(0) = 0$ from equation (5). Next, we can follow the proof process in the previous part to get

$$\lambda_{2c} = \frac{\gamma_2}{\langle k \rangle}. \quad (11)$$

So, when $0 \leq \lambda_1 \leq \lambda_{1c}$, the rumor must not spread widely in the generated network.

3.3. Steady-State Analysis of 2SIH2R Model. In this part, we consider a relatively general situation. At the beginning of model spreading, in this situation, we assumed that there is one spreader1 who spreads the rumor, and one spreader2 who spreads the truth. So, the initial conditions can be given: $I(0) = (N - 2/N) \approx 1$, $S_1(0) = S_2(0) = (1/N) \approx 0$, $H(0) = R_1(0) = R_2(0) = 0$.

Moreover, it is worth noting that in this situation, due to the complicated spreading process, we can not follow the proof process in the previous part to get the condition of 2SIH2R model spreading threshold. Therefore, this part restarts from the initial conditions and gets the condition of 2SIH2R model spreading threshold.

We use $i(t)$, $s_1(t)$, $s_2(t)$, $h(t)$, $r_1(t)$, $r_2(t)$, separately, to represent from equations (2) to (7), and from the initial conditions we can know

$$i(0) = -2\langle k \rangle \frac{(N - 2)}{N^2}, \quad (12)$$

$$s_1(0) = (1 - f(m))\lambda_1 \langle k \rangle \frac{N - 2}{N^2} - (\alpha + \beta_1) \langle k \rangle \frac{1}{N^2} - \frac{\gamma_1}{N}, \quad (13)$$

$$s_2(0) = \lambda_2 \langle k \rangle \frac{N - 2}{N^2} - \beta_2 \langle k \rangle \frac{1}{N^2} - \frac{\gamma_2}{N}, \quad (14)$$

$$h(0) = f(m)\eta\langle k\rangle \frac{N-2}{N^2}, \quad (15)$$

$$r_1(0) = (1 - (1 - f(m))\lambda_1 - f(m)\eta)\langle k\rangle \frac{N-2}{N^2} + \beta_1\langle k\rangle \frac{1}{N^2} + \frac{\gamma_1}{N}, \quad (16)$$

$$r_2(0) = (1 - \lambda_2)\langle k\rangle \frac{N-2}{N^2} + \alpha\langle k\rangle \frac{1}{N^2} + \beta_2\langle k\rangle \frac{1}{N^2} + \frac{\gamma_2}{N}. \quad (17)$$

Here, $i(0)$ represents the instantaneous rate of change of the ignorant when $t = 0$. The $r(0) = r_1(0) + r_2(0)$ represents the instantaneous rate of change of the stifler1 and stifler2. If the 2SIH2R model can work successfully, at $t = 0$, some ignorant people must become the other five categories. So, the number of ignorant people decreases, and the number of the other five categories increases, which means $i(0) < 0$, $|i(0)| > |r(0)|$ and $r(0) \geq 0$.

From equations (12) to (17), we have

$$i(0) + s_1(0) + s_2(0) + h(0) + r_1(0) + r_2(0) = 0. \quad (18)$$

Since $r(0) = r_1(0) + r_2(0) \geq 0$, we have

$$s_1(0) + s_2(0) + h(0) > 0. \quad (19)$$

So,

$$\begin{aligned} s_1(0) + s_2(0) + h(0) &= (1 - f(m))\lambda_1\langle k\rangle \frac{N-2}{N^2} \\ &\quad - (\alpha + \beta_1)\langle k\rangle \frac{1}{N^2} - \frac{\gamma_1}{N} \\ &\quad + \lambda_2\langle k\rangle \frac{N-2}{N^2} - \beta_2\langle k\rangle \frac{1}{N^2} - \frac{\gamma_2}{N} \\ &\quad + f(m)\eta\langle k\rangle \frac{N-2}{N^2} > 0. \end{aligned} \quad (20)$$

From equation (20), we have

$$(1 - f(m))\lambda_1 + \lambda_2 + f(m)\eta > \frac{\alpha + \beta_1 + \beta_2}{N-2} + \frac{N(\gamma_1 + \gamma_2)}{(N-2)\langle k\rangle}. \quad (21)$$

When $N \rightarrow \infty$, the following result can be obtained:

$$(1 - f(m))\lambda_1 + \lambda_2 > \frac{\gamma_1}{\langle k\rangle} - f(m)\eta + \frac{\gamma_2}{\langle k\rangle}. \quad (22)$$

So, if the rumor and the truth can spread widely in the generated network which is closed and homogeneous, the λ_1 and λ_2 should satisfy equation (22). Next, setting $\lambda_2 = 0, \gamma_2 = 0$, we can get equation (10). And by setting $\lambda_1 = 0, f(m) = 0$, we can get equation (11). So, we can conclude that the third general situation contains the first two special situations.

4. Numerical Simulation

In this section, through numerical simulation, we study the influence of discernible mechanism, confrontation mechanism, and average degree on the rumor. According to the 2SIH2R rumor spreading model and existing research results [29–31], we perform numerical simulation in a generated homogeneous network, where $\langle k\rangle = 8$, $N = 10^5$. It is assumed that there are one spreader1 and one spreader2 at the time $t = 0$. So, $I(0) = (N - 2)/N \approx 1$, $S_1(0) = S_2(0) = (1/N) \approx 0$, $H(0) = R_1(0) = R_2(0) = 0$.

Figure 2 displays the change of density of six categories (spreader1, spreader2, stifler1, stifler2, ignorant, hesitant) over time with $f(m) = 0.7, m = 0.3, \lambda_1 = \lambda_2 = 0.7, \eta = 0.8, \theta_1 = 0.5, \theta_2 = 0.3, \beta_1 = \beta_2 = 0.3, \gamma_1 = \gamma_2 = 0.1, \omega = 0, \alpha = 0.5$. Unless otherwise specified, the above parameters are used in this section. It can be seen from Figure 2 that the density of the ignorant decreases rapidly and those of the other 5 categories increase to their peak, separately in a short time. As the model spreads further, the densities of spreader1 and spreader2 will continue decreasing until they reach zero, which means the 2SIH2R model gets into the steady state.

Figure 3 displays the change of density of spreader1, under the change of parameter m . It can be seen that the greater the m (the bigger the $f(m)$), the stronger the discernible mechanism, the smaller the impact of the rumor, because of the decreasing peak. At the same time, from Figure 4, as m increases, the final size of stifler1 also decreases. But the time to peak of spreader1 and stifler1 has not changed significantly. In Figure 5, it also can be seen that the final size of stifler1 decreases with increasing $f(m)$, but the stifler2 increases. In summary, as m increases, the instantaneous maximum influence and the final influence range of the rumor will decrease but the truth increases.

Figure 6 displays the change of density of spreader1, under the change of parameter α . It can be seen that the greater the α , the stronger the confrontation mechanism, the smaller the impact of the rumor, because of the decreasing peak. At the same time, from Figure 7, as α increases, the final size of stifler2 increases, because some spreader1 change into stifler2 by the confrontation mechanism. In Figure 8, it can also be seen that the final size of stifler1 decreases with increased α , but the stifler2 increases. In summary, as α increases, the instantaneous maximum influence and the final influence range of the rumor will decrease, while the final influence range of the truth will increase.

Figure 9 displays the change of density of spreader1, under the change of parameter $\langle k\rangle$. It can be seen that the greater the $\langle k\rangle$, the more people can be contacted by spreader1, the greater the impact of the rumor, because of the increased peak and the shortened time of reaching the peak. Moreover, we can find that, with increased $\langle k\rangle$, the shape of the solid line becomes wider, which means that the duration of the rumor event is decreasing. In summary, as $\langle k\rangle$ increases, the velocity and the range of the rumor spreading will increase, which means the influence of the

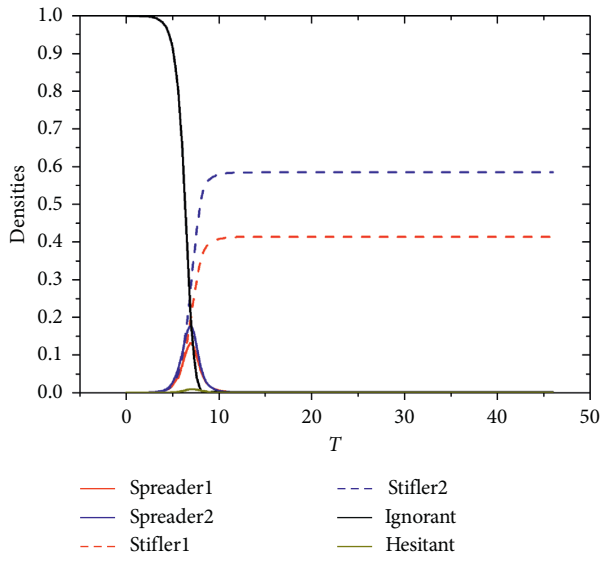


FIGURE 2: Density of 6 categories of people in homogeneous network.

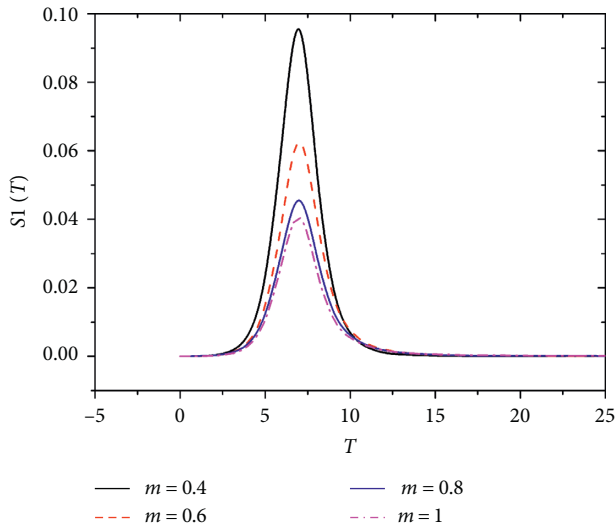


FIGURE 3: Density of spreader1 over time for different values of m .

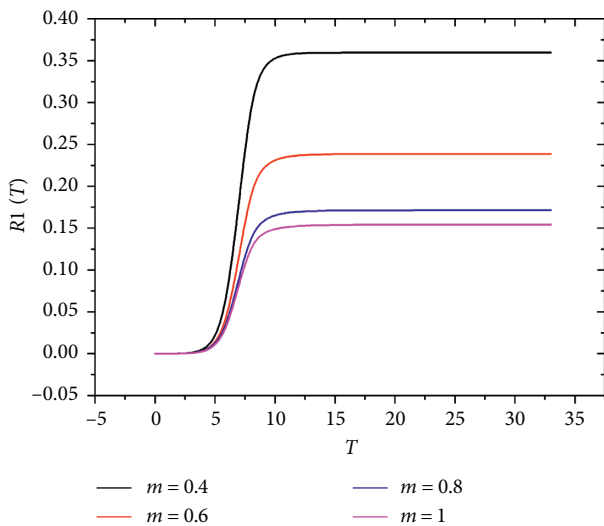


FIGURE 4: Density of stifler1 over time for different values of m .

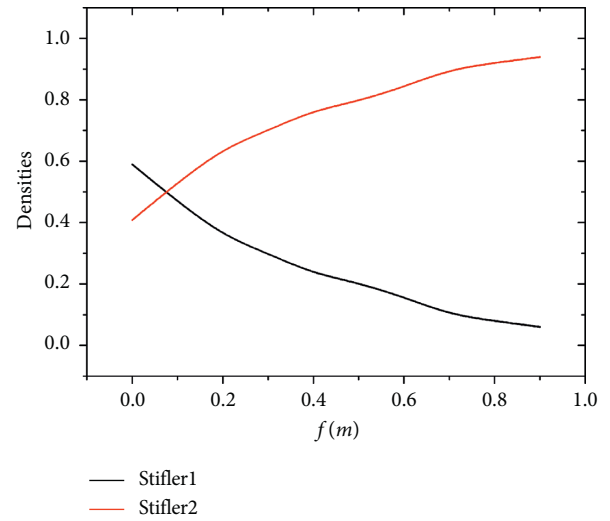


FIGURE 5: Density of stifler1 and stifler2 for different values of $f(m)$ at steady state.

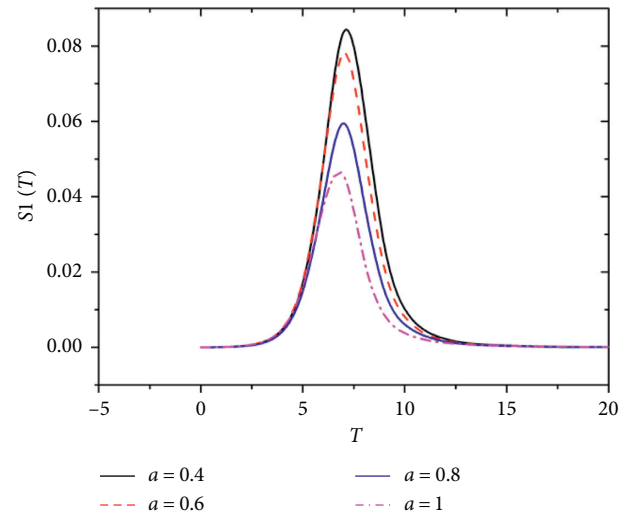


FIGURE 6: Density of spreader1 over time for different values of α .

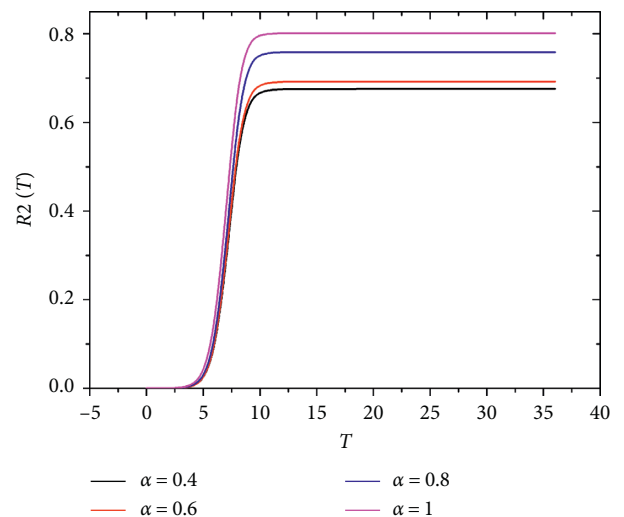


FIGURE 7: Density of stifler2 over time for different values of α .

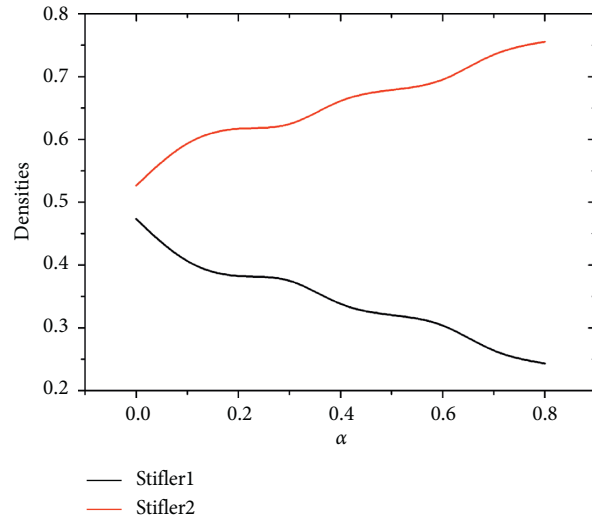


FIGURE 8: Density of stifler1 and stifler2 for different values of α at steady state.

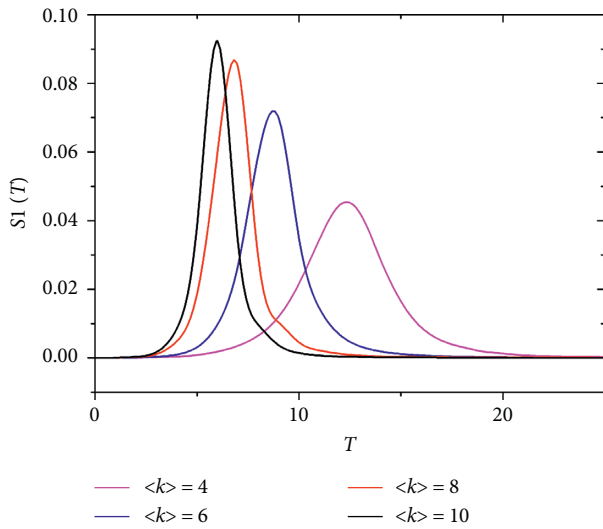


FIGURE 9: Density of spreader1 over time for different values of $\langle k \rangle$.

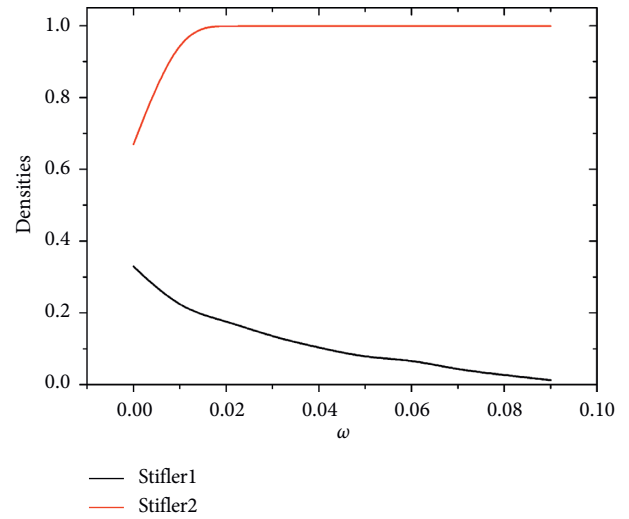


FIGURE 10: Density of stifler1 and stifler2 for different values of ω at steady state.

rumor will increase significantly. But the duration of the rumor event will decrease.

Figure 10 displays the change of density of stifler1 and stifler2, under the change of parameter ω . It can be seen that the change of ω causes a huge impact on the rumor. As long as ω changes from 0 to 0.1, almost only stifler2 exists in the network when it reaches a steady state, which means the rumors will not have a significant impact on us. So, this paper mainly studies the situation where $\omega = 0$.

Figures 11 and 12 display the final size R (which is the sum densities of stifler1 and stifler2 at steady state) with λ_1 and λ_2 . The redder the color is, the greater the value of R . In Figure 12, under the parameter $f(m) = 0.5$, $\eta = 0.1$, $\gamma_1 = \gamma_2 = 0.8$, the spreading threshold condition can be distinguished roughly by the shade of color (as the black solid line denoted in the figure) which is basically consistent with the steady-state analysis from the previous section (as the black dashed line denoted in the figure).

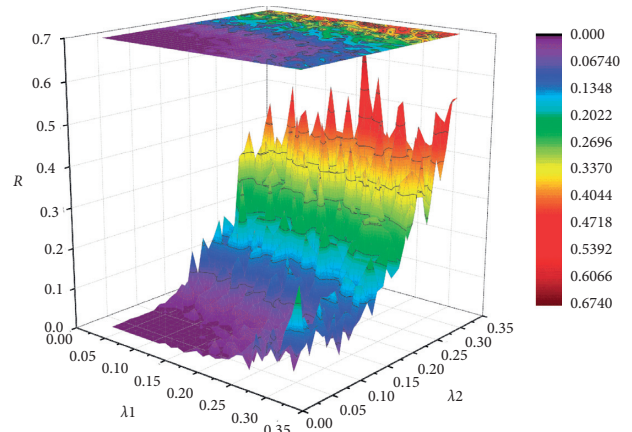


FIGURE 11: Density of R with λ_1 and λ_2 under the parameter $f(m) = 0.5$, $\eta = 0.1$, $\gamma_1 = \gamma_2 = 0.8$.

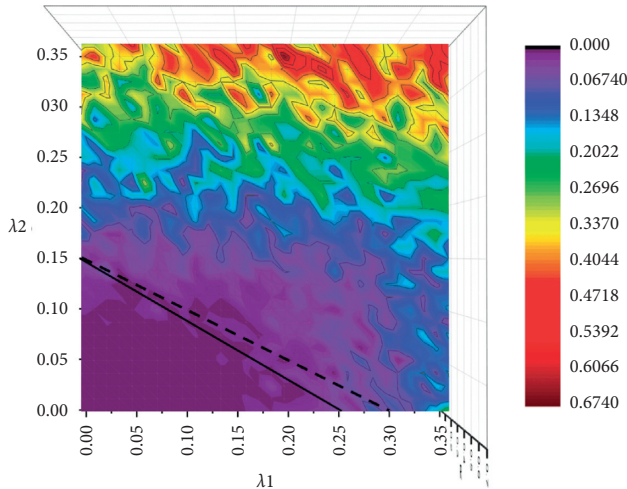


FIGURE 12: The vertical view of Figure 11.

5. Conclusions

Rumors as a kind of harmful information in most situation may cause large public panic. It is necessary to establish a suitable mathematical model to analyse the characteristics and mechanisms of the rumor dissemination process. In this paper, we propose a $2SIH2R$ rumor spreading model in a generated homogeneous network, and some significant results can be given:

- (1) We assumed that when an ignorant encounters a spreader1, the ignorant may change into hesitant1 because of the discernible mechanism, and when a spreader1 encounters a spreader2, the spreader1 may change into stifier2 because of the confrontation mechanism.
- (2) Through changing parameters, the model can be simplified to the traditional SIR model and the SEIR model. From this point, the traditional model has been improved and the $2SIH2R$ model is more universal.
- (3) After establishing the mean-field equations of the $2SIH2R$ model, we give the condition of $2SIH2R$ model spreading threshold in three situations separately. When there is only one spreader1 at $t = 0$ in the networks, the spreading threshold is $\lambda_{1c} = (\gamma_1 - \langle k \rangle f(m)\eta / \langle k \rangle (1 - f(m)))$. When there is only one spreader2 at $t = 0$ in the networks, the spreading threshold is $\lambda_{2c} = (\gamma_2 / \langle k \rangle)$. When there is one spreader1 and one spreader2 at $t = 0$, the condition of $2SIH2R$ model spreading threshold is $(1 - f(m))\lambda_1 + \lambda_2 > (\gamma_1 + \gamma_2 / \langle k \rangle) - f(m)\eta$. When the condition is not satisfied, the rumor or the truth cannot spread widely in the crowd.
- (4) From the numerical simulations, we can know that the higher the discernible degree is, the smaller the influence of the rumor will be; the higher the confrontation rate is, the smaller the influence of the rumor will be; the bigger the average degree is, the

greater the influence of the rumor will be, but the shorter the duration is.

In the future, a further study on $2SIH2R$ rumor spreading model will be conducted in the heterogeneous and some real networks. In this paper, we assume that the social network is homogeneous, but in reality, lots of social networks have a more complex structure. And also, the real data may be analysed, because there are many subjective assumptions in the model and parameters setting process. The significance of the model can be better demonstrated through real data.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This paper was financially supported by Beijing Municipal Natural Science Foundation (9202018), Fundamental Research Funds for the Central Universities (CUC2000004, CUC19ZD002, and 2018CUCTJ047), and Academic Research Project of China Federation of Radio and Television Association (2020ZGLH015).

References

- [1] J. P. Zhang, H.-M. Guo, J. Wen-Jun, and Z. Jin, "Dynamic analysis of rumor propagation model based on true information spreader," *Acta Physica Sinica*, vol. 68, no. 15, pp. 193–204, 2019.
- [2] L.-H. Zhu and Li Ling, "Dynamic analysis of rumor-spread-delaying model based on rumor-refuting mechanism," *Acta Physica Sinica*, vol. 69, no. 2, pp. 67–77, 2020.
- [3] A. Amankwah and A.-A. Joseph, "Integrating the dark side of competition into explanations of business failures: evidence from a developing economy," *European Management Review*, vol. 15, no. 1, 109 pages, 2017.
- [4] Y. Moreno, M. Nekovee, and A. F. Pacheco, "Dynamics of rumor spreading in complex networks," *Physical Review E*, vol. 69, no. 6, Article ID 066130, 2004.
- [5] D. J. Daley and D. G. Kendall, "Epidemics and rumours," *Nature*, vol. 204, no. 1118, 1964.
- [6] D. H. Zanette, "Critical behavior of propagation on small-world networks," *Physical Review E*, vol. 64, Article ID 050901, 2001.
- [7] D. H. Zanette, "Dynamics of rumor propagation on small-world networks," *Physical Review E*, vol. 65, no. 4, Article ID 041908, 2002.
- [8] Yi-P. Wan, D.-Ge Zhang, and Q.-H. Ren, "Propagation and inhibition of online rumor with considering rumor elimination process," *Acta Physica Sinica*, vol. 64, no. 24, Article ID 240501, 2015.
- [9] M. Zhao, W. Chen, and Q. Song, "Research on a rumor spreading model with media coverage," *Applied Mathematics and Mechanics*, vol. 39, no. 12, pp. 1400–1409, 2018.

- [10] M. Askarizadeha and M. Hossein Manshaei, "An evolutionary game model for analysis of rumor propagation and control in social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 523, no. 1, pp. 21–39, 2019.
- [11] Y. Zhang and J. Zhu, "Stability analysis of I2S2R rumor spreading model in complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 503, no. 1, pp. 862–881, 2018.
- [12] L. a. Huo, F. Ding, and Y. Cheng, "Dynamic analysis of a SIbInIu, rumor spreading model in complex social network," *Physica A: Statistical Mechanics and Its Applications*, vol. 523, no. 1, pp. 924–932, 2019.
- [13] S. Deng and W. Li, "Spreading dynamics of forget-remember mechanism," *Physical Review E*, vol. 95, no. 4, Article ID 042306, 2017.
- [14] Yi-R. Gu and L.-L. Xia, "Propagation and inhibition of rumors in online social network," *Acta Physica Sinica*, vol. 61, no. 23, Article ID 238701, 2012.
- [15] J. Wang, L. Zhao, and R. Huang, "SIRaRu rumor spreading model in complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 398, no. 15, pp. 43–55, 2014.
- [16] J. Wang, L. Zhao, and R. Huang, "2SI2R rumor spreading model in homogeneous networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 431, no. 1, pp. 153–161, 2014.
- [17] T. Wang, J. He, and X. Wang, "An information spreading model based on online social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 490, no. 15, pp. 488–496, 2018.
- [18] A. Yang, X. Huang, X. Cai, X. Zhu, and L. Lu, "ILSR rumor spreading model with degree in complex network," *Physica A: Statistical Mechanics and Its Applications*, vol. 531, no. 1, Article ID 121807, 2019.
- [19] L.-L. Xia, G.-P. Jiang, B. Song, and Y. R. Song, "Rumor spreading model considering hesitating mechanism in complex social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 437, no. 1, pp. 295–303, 2015.
- [20] L. Huo, L. Wang, N. Song, C. Ma, and B. He, "Rumor spreading model considering the activity of spreaders in the homogeneous network," *Physica A: Statistical Mechanics and Its Applications*, vol. 468, no. 15, pp. 855–865, 2017.
- [21] L. Huo and Y. Cheng, "Dynamical analysis of a IWSR rumor spreading model with considering the self-growth mechanism and indiscernible degree," *Physica A: Statistical Mechanics and Its Applications*, vol. 536, no. 15, Article ID 120940, 2019.
- [22] M. Ran, C. Liu, X. Huang et al., "Rumor spread model considering difference of individual interest degree and refutation mechanism," *Journal of Computer Applications*, vol. 38, no. 11, pp. 3312–3318, 2018.
- [23] X. Dong, Y. Liu, C. Wu, Y. Lian, and D. Tang, "A double-identity rumor spreading model," *Physica A: Statistical Mechanics and Its Applications*, vol. 528, no. 15, Article ID 121479, 2019.
- [24] Y. Zan, J. Wu, P. Li, and Q. Yu, "SICR rumor spreading model in complex networks: counterattack and self-resistance," *Physica A: Statistical Mechanics and Its Applications*, vol. 405, no. 1, pp. 159–170, 2014.
- [25] G. W. Allport and L. Postman, *The Psychology of Rumor*, 1947.
- [26] J. Li, H. Jiang, Z. Yu, and C. Hu, "Dynamical analysis of rumor spreading model in homogeneous complex networks," *Applied Mathematics and Computation*, vol. 359, no. 15, pp. 374–385, 2019.
- [27] X. Rui, F. Meng, Z. Wang, G. Yuan, and C. Du, "SPIR: the potential spreaders involved SIR model for information diffusion in social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 506, no. 15, pp. 254–269, 2018.
- [28] M. Kuperman and G. Abramson, "Small world effect in an epidemiological model," *Physical Review Letters*, vol. 86, no. 13, Article ID 2909, 2001.
- [29] A. Hosni, Li Kan, and S. Ahmad, "Analysis of the impact of online social networks addiction on the propagation of rumors," *Physica A: Statistical Mechanics and Its Applications*, vol. 542, no. 15, Article ID 123456, 2020.
- [30] D. Fan, G.-P. Jiang, Y.-R. Song, and Y.-W. Li, "Novel fake news spreading model with similarity on PSO-based networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 549, no. 1, Article ID 124319, 2020.
- [31] C.-Y. Sang and S.-G. Liao, "Modeling and simulation of information dissemination model considering user's awareness behavior in mobile social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 537, no. 1, Article ID 122639, 2020.

Research Article

Evolving Nature of Human Contact Networks with Its Impact on Epidemic Processes

Cong Li ^{1,2,3}, Jing Li ¹ and Xiang Li ^{1,2,3}

¹Adaptive Networks and Control Lab, Department of Electronic Engineering, Fudan University, Fudan 200433, China

²Research Center of Smart Networks and Systems, School of Information Science and Engineering, Fudan University, Fudan 200433, China

³MOE Frontiers Center for Brain Science, Fudan University, Fudan 200433, China

Correspondence should be addressed to Xiang Li; lix@fudan.edu.cn

Received 1 November 2020; Accepted 16 January 2021; Published 2 February 2021

Academic Editor: Dan Selisteanu

Copyright © 2021 Cong Li 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.

Human contact networks constitute a multitude of individuals and pairwise contacts among them. However, the dynamic nature, which generates the evolution of human contact networks, of contact patterns is unknown yet. Here, we analyse three empirical datasets and identify two crucial mechanisms of the evolution of temporal human contact networks, i.e., the activity state transition laws for an individual to be socially active and the contact establishment mechanism that active individuals adopt. We consider both of the two mechanisms to propose a temporal network model, the so-called memory-driven (MD) model, of human contact networks. Then, we study the susceptible-infected (SI) spreading processes on empirical human contact networks as well as four corresponding temporal network models and compare the full prevalence time of SI processes with various infection rates on the networks. The full prevalence time of SI processes in the MD model is the same as that in real-world human contact networks. Moreover, we find that the individual activity state transition promotes the spreading process, while the contact establishment of active individuals suppresses the prevalence. Besides, we observe that individuals who establish social ties with a small exploring rate are still able to induce an endemic which prevails in the networks. The study offers new insights to predict and control the diffusion processes on networks.

1. Introduction

The evolution nature of human interactions creates diverse temporal properties, which fundamentally influence the epidemic spreading. Interactions may be characterized and modeled by the human interaction networks [1–5], where a node represents a person, and the interaction between two persons is a link. In 2004, Pentland et al. from the MIT Media Laboratory took the lead in tracking and recording human interactions with personal mobile phones and proposed a stochastic process-based model to capture the coevolution of social relationships and individual behaviors [6, 7]. Many previous studies of interaction networks assumed that the interactions are fixed, or the spreading processes are much faster than the evolution of networks. Nowadays, with the rapid technological advances, high

quality and time-resolved datasets of human interactive behaviors can be easily obtained [8–13] and have provided an unprecedented opportunity to better understand human society dynamics with temporal human interaction networks [14–16]. However, it remains an important challenge to unveil the mechanisms driving the evolution of human interaction networks and to further capture the effects of such mechanisms on dynamical processes on networks [17–20].

In human interaction networks, in general, each agent at time t is either active, contacting with other agents, or inactive, isolated. The activity state, i.e., activeness or inactiveness of an agent and the contacts between pairwise agents alter with time. Rocha and Blondel [21] introduced a temporal network model, where the time interval between two activity states of each agent follows a stochastic process

with a certain distribution, and the inactive agents keep isolated from other active ones in the network. Nevertheless, more research efforts assumed that the nodes in the inactive state are not totally isolated and can receive contacts from other active nodes. A typical example is the framework of activity-driven networks [22–25]. Perra et al. [22] proposed the original activity-driven model of temporal networks, in which an agent is either active with probability p or inactive with probability $1 - p$ at each unit time. Since the active probability of an agent is decided by its activity potential which follows a given distribution, regardless of its activity state at the last time step, the activity-driven model generates a sequence of memoryless networks. Besides, the social contacts between active nodes and other nodes are randomly established. An extension of this model reduced the randomness of link establishments by introducing a mutual selection mechanism, where the destination of a contact depends mutually on the activity potential of the agents on both ends of the connection [23]. Similarly, some studies added a time-invariant quantity, namely, attractiveness for each agent, accounting for the fact that some agents are more likely to be connected by other agents when building a contact [24, 25].

A wealth of empirical observations has shown that the mechanisms which govern the evolution of human contact networks are far from random [14–16, 18, 26–29]. Individuals often remember the agents whom they previously interacted with and thus form their own social circles [18]. Besides, due to the continuity of social interactions, the activity state of an agent is unlikely to be altered in a sudden [13], especially when the agent is actively engaging in a social event with a lot of peers. Valdano et al. [28] proposed a homogeneous assumption that the transition probability between the active state and inactive state of each agent is identically assigned with the same value. Although the assumption partially captures the evolution of the activity state, an explanation of the transition mechanism is still missing. When it comes to the evolution of contacts between agents, the effect of memory, i.e., an agent preserving their previous contacts with a certain probability, should be considered within the temporal network model. The model introduced in [19] incorporates four distinct memory mechanisms to describe the rates of creation and disappearance of contacts in empirical networks, where the rates follow a power-law decaying functional forms, akin to the preferential attachment mechanism [30]. Recently, more realistic mechanisms based on the activity-driven framework have been explored [14, 15, 26, 27]. They considered a non-Markovian reinforcement process, in which agents are more inclined to allocate their contacts towards already existing social ties rather than create new relationships. In particular, the probability of an agent exploring a new social relationship is determined by its cumulative degree, i.e., the number of distinct agents contacted during observation period; however, the transition of the activity states of each agent is not taken into account.

The aforementioned studies have investigated various topological and temporal features of real social systems from different perspectives. However, a better understanding of

the mechanisms driving the evolution of the temporal social contact networks and the effects of network evolution on dynamic processes, such as epidemic spreading [31–34], information propagation [20, 35], and innovation diffusion [36, 37], is still urgently needed.

Exploring several datasets of human contact networks, we report two generic mechanisms of the evolution of real networks: (i) the transition probability of activity states is a function of the degree of agents and (ii) the time interval between the reconnection of node pairs follows a power-law distribution. Based on the two mechanisms, we propose a temporal network model, namely, the memory-driven model, to characterize the evolution of social networks. The memory here refers that an agent inclines to concentrate towards recently contacted partners. With the new framework, we study the epidemic spreading processes in both synthetic and real temporal networks. The full prevalence time of susceptible-infected (SI) spreading in the memory-driven model is the same as that in real networks. Moreover, we find that the two mechanisms of the network evolution have distinct effects on the epidemic spreading on real networks; that is, the activity state transition promotes the spreading process; however, the contact memory hampers the spreading. Apart from this, we demonstrate that even a smaller fraction of interactions connecting to new social relationships is able to induce an explosive spreading on networks.

This paper is organized as follows. In Section 2, we describe three datasets of real-world human interactive activities and introduce the definitions of human contact networks. In Section 3, we analyse the empirical datasets to obtain statistic properties of human contact networks and propose a temporal network model. Section 4 focuses on the epidemic spreading processes on the synthetic networks and real networks. Finally, we conclude this work in Section 5.

2. Empirical Datasets and Definitions

In this paper, we use three empirical datasets of offline human interactive activities in various scenes. The datasets include the offline sex interaction dataset (“Sex6yr,” [38, 39]) and the physical proximity interaction datasets (“MIT_RM,” [6] and “School,” [40]). The “Sex6yr” dataset collects the real offline sex trading activities among sellers and buyers with a forum-like Brazilian web community. If a buyer A posted a comment in a thread about seller B, a contact between A and B exists. The “MIT_RM” dataset records the call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and so on among students and faculty in the MIT Media Laboratory via the Bluetooth device in mobile phones. We here studied the physical proximity interactions with the sampling data of physical proximity via Bluetooth devices in the “MIT_RM” dataset. The “School” dataset collects the face-to-face proximity interactions among teachers and students in a French school by the Radio-Frequency Identification (RFID) device embedded in badges. The contact records of each dataset are in the same format (i, j, t) , which represents a contact between two individuals i and j started at time t . More details of the

datasets are introduced in Appendix A and summarized in Table 1.

The human contact network is a kind of temporal network, which consists of the evolution of network topology with time. We segment the empirical data into adjacency time steps of length $t_w = w\Delta t$, where Δt is the resolution and w is the number of resolutions in a time step. A human contact network is a sequence of networks in time order $\mathcal{G}(t_w) = \{G_1, G_2, \dots, G_t, \dots, G_T\}$, where $G_t = (\mathcal{N}_t, \mathcal{L}_t)$ is the network taking place at time step t over time window $[(t-1)t_w, t \cdot t_w)$ and T is the total number of time steps. An illustration of the human contact network is shown in Figure 1. In this work, we select t_w as 1 month, 1 day, and 5 mins for “Sex6yr,” “MIT_RM,” and “School,” respectively (explanation for the selection is given in Appendix B). The aggregated network of $\mathcal{G}(t_w)$ is $\mathcal{G}_T = \cup_{t=1}^{t=T} G_t = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N} = \cup_{t=1}^{t=T} \mathcal{N}_t$ and $\mathcal{L} = \cup_{t=1}^{t=T} \mathcal{L}_t$. We denote the number of agents and the number of contacts by $N = |\mathcal{N}|$ and $L_C = |\mathcal{L}|$, respectively.

3. Analysis of Temporal Properties

Before we introduce the temporal network model, we first empirically analyse statistic properties of human contact networks, i.e., the activity state transition of individuals and the contact memory of node pairs, which are the fundamental mechanisms for the design of our temporal network model.

3.1. Individual Activity State Transition. We here assume that each node at time step t in the network is in one of two states: active state, connecting with other nodes, denoted by a and inactive state i . The sets of nodes in active state and in inactive state at time step t are denoted by $\mathcal{N}_a(t)$ and $\mathcal{N}_i(t)$, respectively. We here study whether the activity state of an individual j at time step $t+1$ depends on the state at previous step t and how the degree k of individual j at time step t influences the activity state transition. We denote the number of nodes with degree k in active state at time step t by $N_{k,a}(t) = |\mathcal{N}_{k,a}(t)|$, where $|\mathcal{N}|$ represents the number of nodes in a node set \mathcal{N} . The transition probability of nodes with degree k transferring from active state a to inactive state i at one time step is

$$P_{AI}(k) = \frac{\sum_{t=1}^{T-1} |\mathcal{N}_{k,a}(t) \cap \mathcal{N}_i(t+1)|}{\sum_{t=1}^{T-1} N_{k,a}(t)}. \quad (1)$$

The transition probability of nodes with degree k staying at active state a in one time step is $P_{AA}(k) = 1 - P_{AI}(k)$. Note that the degree of an individual might change over time since the individual might take part in or leave from the interactions of the system across time.

The transition probabilities characterize the relation between the activity state of node i with degree k at the current time step and that at next time step. Although it has been observed in empirical networks (e.g., Facebook) that the strong link between activity and degree [18], the detail

TABLE 1: Description of the three datasets.

| Dataset | T | Δt | N | L_C | R |
|---------|-----------|------------|--------|--------|--------|
| Sex6yr | 2,232 day | 1 day | 16,102 | 38,995 | 50,185 |
| MIT_RM | 232 day | 6 min | 96 | 2,539 | 55,306 |
| School | 8.67 hour | 20 sec | 236 | 5,901 | 37,402 |

T is the time span of contact sequences, Δt is the resolution (time interval between two records in time order), N is the number of agents, L_C is the number of distinct contacts, and R is the number of records.

about how they are related and how to apply it to network modeling is not fully discussed previously. Here, we study how the transition probability, i.e., $P_{AI}(k)$ or $P_{AA}(k)$, varies with the degree k . Figures 2(a)–2(c) show the transition probabilities of remaining in the active status or switching to the inactive status calculated on three empirical networks. Remarkably, the transition probability $P_{AI}(k)$ or $P_{AA}(k)$ is as a power-law function of degree k . The greater the degree of active agents at the current time step, the smaller the probability they become inactive at the next time step. That is to say, agents with high social activity (high degree) have more tendency to maintain the active status all the time. However, we cannot find the relation between the transition probability $P_{IA}(k)$ (or $P_{II}(k)$) and degree k since the degree of nodes in inactive state is 0. Hence, we study the transition property of inactive nodes with the firing rate $b = \mathcal{N}_i(t) \cap \mathcal{N}_a(t+1) / N_i(t)$, which is the probability of an inactive node becoming active at one time step. The average firing rate P_{IA} is expressed as

$$P_{IA} = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{|\mathcal{N}_i(t) \cap \mathcal{N}_a(t+1)|}{N_i(t)}. \quad (2)$$

Figure 2(d) demonstrates that the variance of firing rates at different time steps is small.

3.2. Contact Establishment of Active Individuals. Previous studies [22–25] assume that all active individuals contact with s other individuals at each time step, where $s = m$ is a constant. However, we find that the number of contacts (degree) at each time step follows a different distribution in empirical datasets (see Figure 3(a)). Figure 3 suggests that the assumption of the same number of contacts cannot reflect the real properties; thus, the distribution $F(s)$ of contact number s should be considered for the temporal network models with different datasets.

An active individual either establishes a contact by social keeping, connecting to a node which has already been connected to at previous time steps, or social exploring, connecting to a new one. The average fraction of social keeping contacts of each individual at each time step is

$$p = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T p_i(t), \quad (3)$$

in which the fraction $p_i(t)$ of reconnected links of node i is

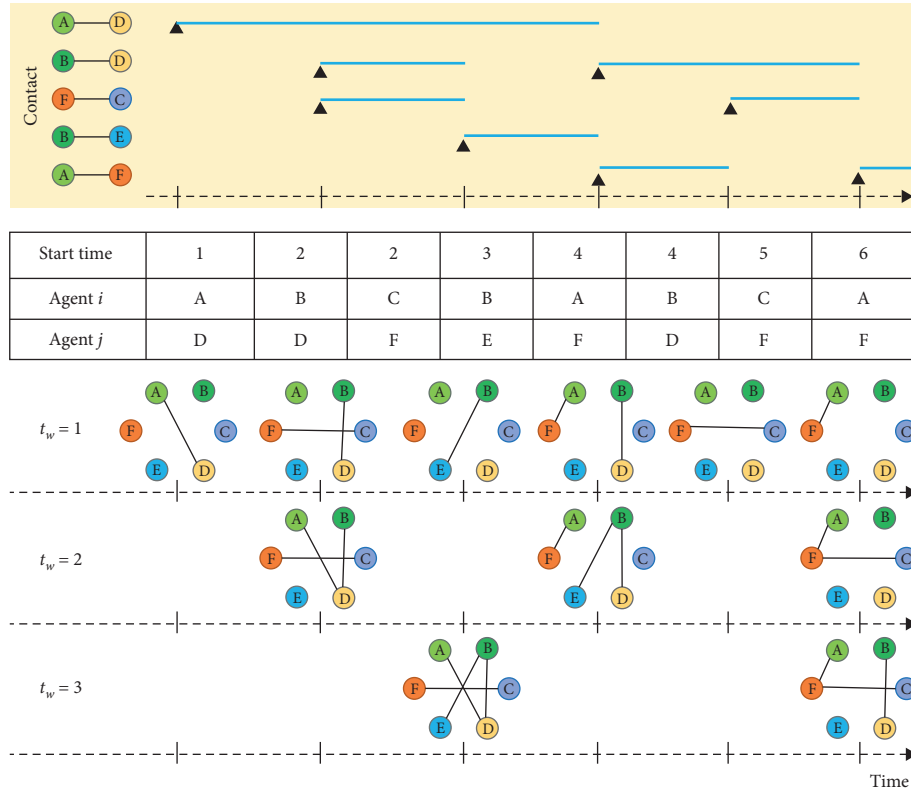


FIGURE 1: Schematic illustration of the construction of temporal networks and the time slice of a contact sequence, where t_w is the length of a time step.

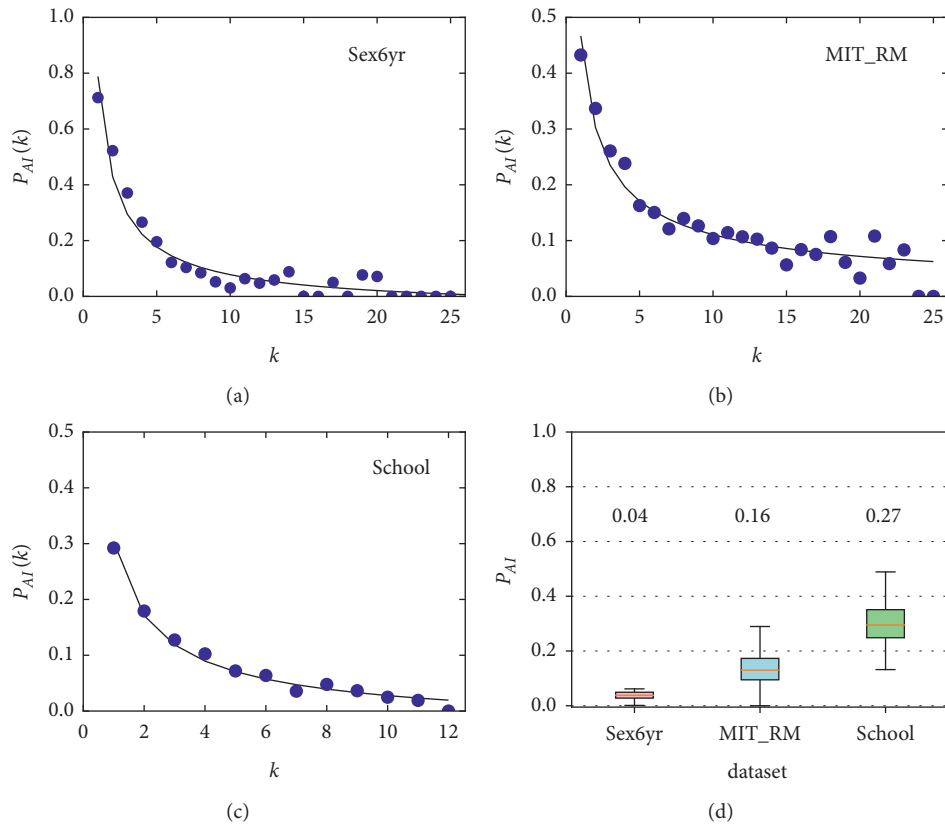


FIGURE 2: Transition probabilities between active state and inactive state as functions of the current degree k of the node, for the three empirical datasets. Solid lines represent the fitting curves: (a) $P_{AI}(k) = 0.78k^{-0.84}$ (Sex6yr), (b) $P_{AI}(k) = 0.45k^{-0.56}$ (MIT_RM), and (c) $P_{AI}(k) = 0.33k^{-0.91}$ (School). (d) The firing rate P_{IA} with 95% confidence intervals in the three empirical datasets. The average firing rates P_{IA} are 0.04 (Sex6yr), 0.16 (MIT_RM), and 0.27 (School).

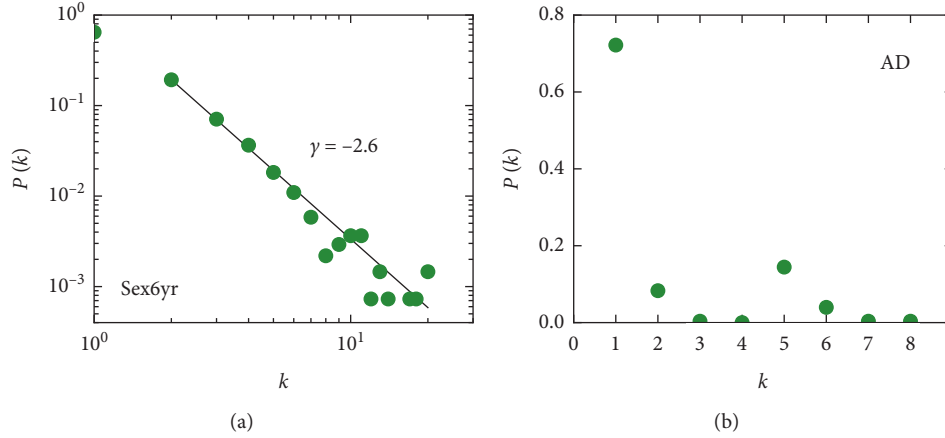


FIGURE 3: (a) Degree distributions $P(k)$ at each time step in an empirical dataset and (b) degree distribution $P(k)$ at each time step in the activity-driven (AD) model [22] (to be described in Section 5.1) with $s = 5$ and $N = 1000$, and node activity x_i is sampled from $F(x_i) \propto x_i^{-\nu}$ with $\nu = 2.1$, $10^{-2} \leq x_i \leq 1$.

$$p_i(t) = \frac{\sum_{j \in \mathcal{Z}_i(t)} \delta_{ij}(t)}{|\mathcal{Z}_i(t)|}, \quad (4)$$

where $\delta_{ij}(t) = 0$ if agent i never interacted with agent j before; otherwise, $\delta_{ij}(t) = 1$. The $|\mathcal{Z}_i(t)|$ is the number of direct neighbors of agent i at time step t . The fraction $p_i(t)$ quantifies the inclination of an agent to keep previous established contacts. Figure 4(a) shows that in the context of sex trade, individuals dominantly adopt the strategy of social exploring, while in the context of physical proximity (MIT_RM) and face-to-face communication (School), individuals tend to reconnect with the nodes who have already contacted by social keeping.

Moreover, the recurrence interval of human contacts is analysed. The recurrence interval $\Delta_{(i,j)}$ of individuals i and j refers to the number of time steps between any two consecutive contacts. We can deduce that, if an active node i has a degree k_i at each time step and k_i contacts are established randomly with other nodes, the distribution of recurrence interval is an exponential distribution as $P(\Delta_{(i,j)}) = (k_i/N)(1 - k_i/N)^{(\Delta_{(i,j)} - 1)}$. Note that this is the basic assumption in previous works [14, 26–28]. However, we find that the probability distributions of recurrence intervals of contacts between node i and all others follow a power-law distribution in three datasets (see Figure 4(b)). The finding might give a support for the burst property of temporal networks that the intercontact time distribution has a power-law form (see Figure 4(c)), which has also been reported recently in [15, 41, 42]. The results imply that agents do not establish contacts randomly with others regardless of their pervious contacts. Besides, we have identified the statistical law in real-world systems that agents preferentially contact individuals who have recently been in interact with [29].

To summarize, two typical characteristics coexist in the process of realistic network evolution, individual activity state transition, and contact establishment. In the next section, we propose a temporal network model based on the mechanisms.

4. Modelling Dynamic Contact Networks

4.1. Memory-Driven (MD) Model. In this section, we present a temporal network model, named the memory-driven (MD) model, of human contact networks. The empirical analysis of individual activity state transition and contact establishment are both the basic mechanisms for the MD model. We consider a set of agents \mathcal{N} in a human contact network \mathcal{G} . For each time step, every agent has two possible activity states: active and inactive. We here assume that each individual has the maximum memory length L time steps; in other words, the contact establishment at current time step is only influenced by the connections at pervious L time steps, which is stored in the memory train $G_{\mathcal{M}} = \{G_{t-L+1}, G_{t-L+2}, \dots, G_{t-1}, G_t\}$. We generate L random networks [43] to initialize the temporal network $\mathcal{G} = \{G_1, G_2, \dots, G_L\}$ and the memory train $G_{\mathcal{M}} = \{G_1, G_2, \dots, G_L\}$. The generation of the temporal network $\mathcal{G} = \{G_1, G_2, \dots, G_T\}$ is illustrated in Figure 5 and described as follows:

- (1) Individual activity state transition: an active individual i with degree k at time step t becomes inactive at time step $t + 1$ with probability $P_{AI}(k) = Ak^{-\alpha}$ ($A \in (0, 1]$), while an inactive agent becomes active with a constant firing rate P_{IA} .
- (2) Contact establishment of active individuals: we assign each active individual with s contacts, where s is extracted from a given probability distribution $F(s)$. The contacts are established by the following steps:
 - (a) Social keeping: With a keeping rate $p \in [0, 1]$, agent i connects to a previous contacted agent j with probability $p_{ij} = f(\Delta_{(i,j)}) \propto \Delta_{(i,j)}^{-\nu}$, ($1 \leq \Delta_{(i,j)} \leq L$), which is a function of the time interval $\Delta_{(i,j)}$ since their last interaction.
 - (b) Social exploring: with an exploring rate $q = 1 - p$, the agent randomly interacts with a new agent who has never been contacted by

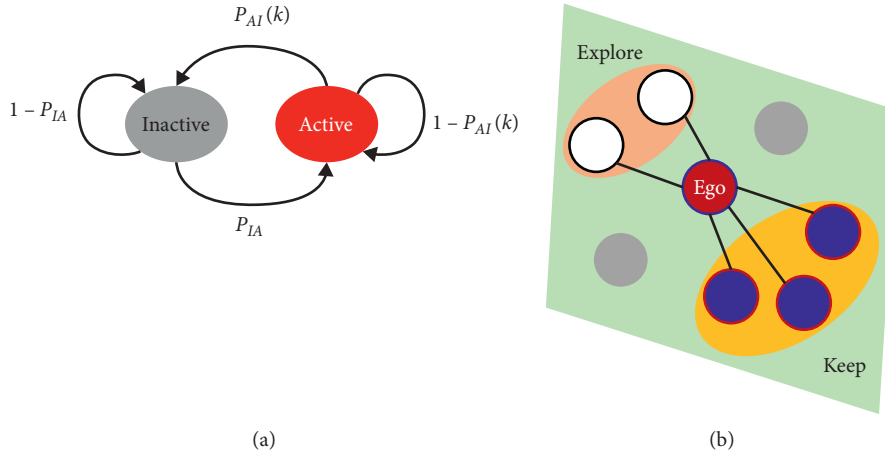
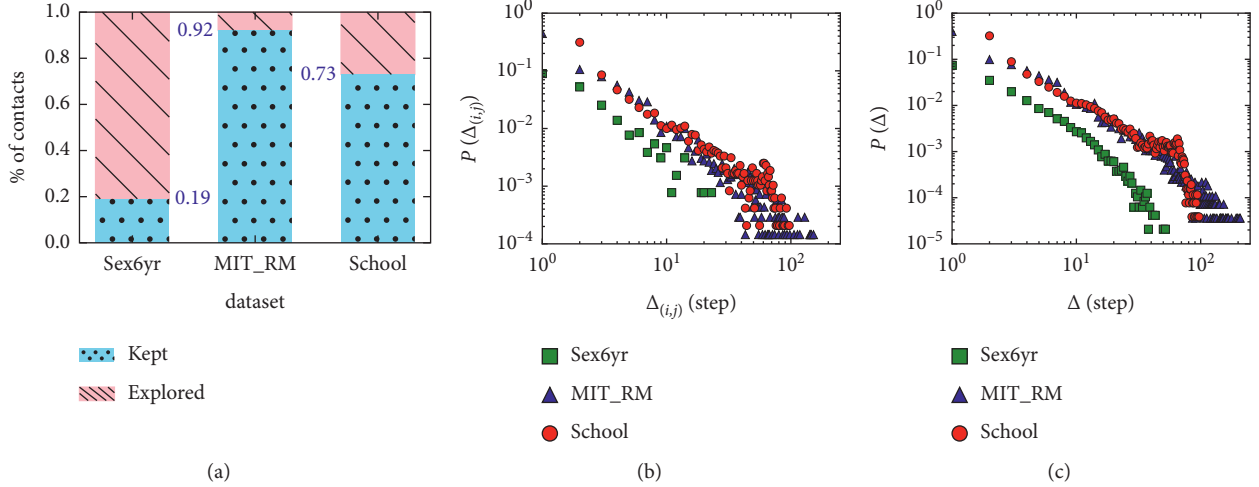


FIGURE 5: (a) Transition probabilities of individual activity states in the MD network model. (b) Contact establishment strategies of each agent in each time step: social keeping (blue) and social exploring (white).

agent i before (or the previous contacts have exceeded the memory length L).

- (3) Generate the network G_{t+1} and update the memory train as $G_{\mathcal{M}} = \{G_{t-L+2}, G_{t-L+3}, \dots, G_t, G_{t+1}\}$.
- (4) Repeat (1)–(3) until the end of time span T of the temporal network.

4.2. Analysis of the Model. We perform numerical simulations in terms of two model settings in Table 2. In Figure 6, the network properties, i.e., the degree distribution ρ_k and the intercontact time distributions $P(\Delta)$, of the memory-driven model are compared with those of the corresponding

TABLE 2: The parameters of the memory-driven model.

| Memory-driven (MD) model | | | |
|--------------------------|------------|-------------------------|-------------------------|
| Parameters | Sex6yr | MIT_RM | School |
| $ \mathcal{N} $ | 16,706 | 96 | 236 |
| T | 64 | 230 | 104 |
| A | 0.78 | 0.45 | 0.33 |
| α | 0.84 | 0.56 | 0.91 |
| P_{IA} | 0.04 | 0.16 | 0.27 |
| $F(s)$ | $Pl(-2.6)$ | $\text{Norm}(9.0, 5.1)$ | $\text{Norm}(7.0, 3.4)$ |
| γ | 1.87 | 1.62 | 1.62 |
| L | 36 | 60 | 12 |

$X \sim Pl(\beta)$ represents the power-law $\rho(x) = (\beta - 1)x^{-\beta}$, where $x \in [1, +\infty)$, and $X \sim \text{Norm}(\mu, \sigma^2)$ represents the Gaussian distribution with the mean μ and variance σ^2 .

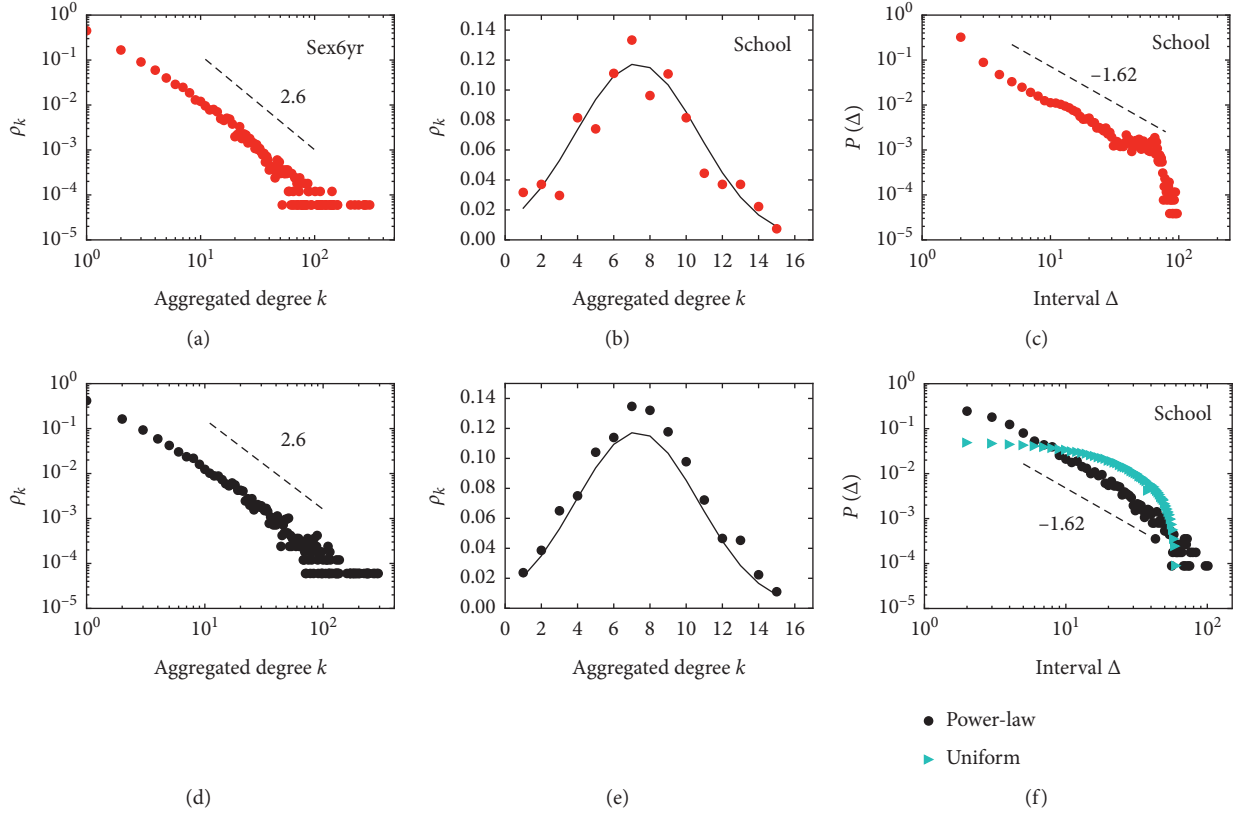


FIGURE 6: Distributions of the characteristic measures of the real-world networks and memory-driven networks. (a, b) The degree distributions ρ_k of the aggregated empirical networks and (d, e) the degree distributions ρ_k of the aggregated memory-driven networks. (c, f) The intercontact time distributions $P(\Delta)$ in the dataset “School,” the memory-driven networks with p_{ij} following a power-law distribution (in black), and the null model with a uniform distribution (in cyan). The synthetic networks are performed with parameters in Table 2.

datasets. The results show that they exhibit pretty similar behaviors; that is, the models reproduce both the degree distributions of the aggregated empirical networks and the intercontact time distributions quite well. Furthermore, we compare the MD model with a null model, whose p_{ij} follows a uniform distribution instead of a power-law distribution. Figure 6(f) indicates a significant difference between the null model and the MD model, implying that the mechanisms of the MD model are responsible for the temporal properties in human contact networks.

Based on the above descriptions, we perform numerical simulations with two different parameter settings, that is, we assume that $F(s)$ follows a power-law distribution and a Gaussian distribution, respectively, to characterize the heterogeneity or homogeneity of human contact networks. The results in Figures 7(a) and 7(b) show that the degree distributions ρ_k of integrated networks generated by our model might be various for different $F(s)$. If $F(s)$ follows a power-law distribution, ρ_k will have a power-law tail with the same exponent γ regardless of the keeping rates p (see Figure 7(a)). If $F(s)$ follows a Gaussian distribution, the functional form of the degree distribution ρ_k does not change with p ; however, the average degree of the integrated network decreases with the increase in p (see Figure 7(b)). The intercontact time distributions $P(\Delta)$ of the two networks are shown in Figures 7(c) and 7(d). When the keeping

rate $p = 0$, the establishment of human contacts is random and memoryless, which leads to an exponential distributions $P(\Delta)$ of the intercontact time (insets of Figures 7(c) and 7(d)). With the increase in p , the intercontact time distribution asymptotically follows a power-law, indicating that the memory embedded in the process of contact establishment can induce the bursty interactive pattern of human activities.

5. Epidemic Processes in Temporal Human Contact Networks and Models

5.1. Dynamical Processes on Real-World Networks and Network Models. We study the susceptible-infected (SI) spreading dynamics [44, 45] on temporal human contact networks. The SI epidemic model, despite of its simplicity, has been pervasively operated as a useful tool to probe the topological structures and temporal characteristics of temporal networks [17, 19, 46]. In this model, each agent can be in two possible states: susceptible (S) or infected (I). A susceptible (S) agent is infected by an infected (I) one with probability λ if there is a contact between them. Once an agent is infected (I), it will not recover. We start each simulation with all agents in the susceptible state and randomly select a single agent as the infected “seed.” In order to uncover the effects of the two mechanisms, i.e., individual

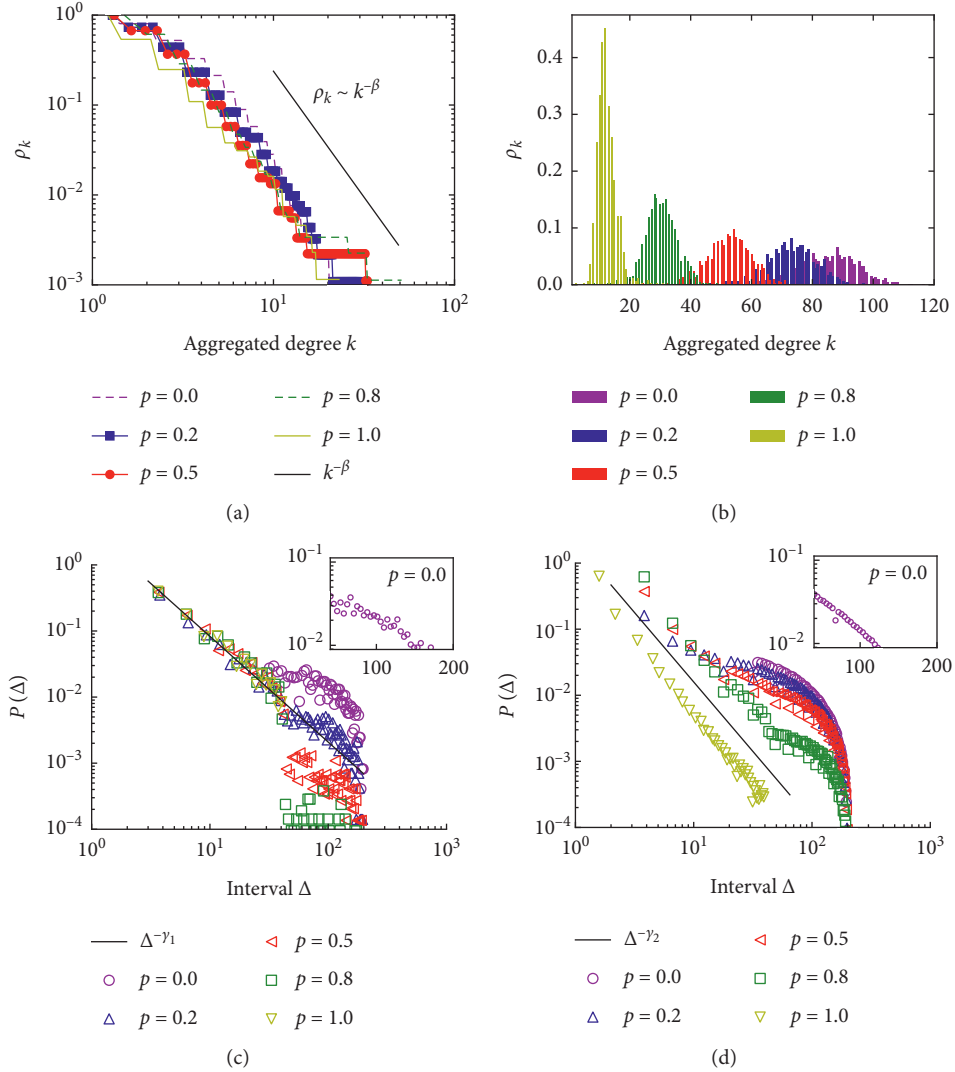


FIGURE 7: (a, b) The degree distributions ρ_k of integrated networks and (c, d) intercontact time distributions $P(\Delta)$ versus the keeping rate p for one simulation. We fix $N = 1000$, $A = 1$, $\alpha = 1$, $P_{IA} = 0.1$, $L = 40$, and $T = 200$. In (a) and (c), the number of contacts s is sampled from a power-law distribution $F(s) = (s/s_{\min})^{-\beta}$ with $\beta = 2.8$, $s_{\min} = 1$, and $\gamma_1 = 1.6$. In (b) and (d), the distribution $F(s)$ satisfies a Gaussian distribution $\text{Norm}(\mu, \sigma^2)$, with the mean $\mu = 5$ and variance $\sigma^2 = 1$ and $\gamma_2 = 1.8$. The intercontact time follows an exponential distribution when $p = 0$ (insets of (c) and (d)).

activity state transition and contact establishment, on the network topologies and further impacts on spreading dynamic processes on temporal networks, we proceed the SI spreading processes on real-world networks and five corresponding network models. The five network models include our memory-driven (MD) model, which has been introduced in Section 4.1, two null models (i.e., AR model and CR model), and two synthetic network models.

We here first introduce two null models where each of the mechanisms is separately destroyed. Besides, as a contrast, we reproduce the temporal network of empirical datasets with two classes of the activity-driven model; that is, activity-driven networks without memory (AD model) and with memory (ADM model) [22, 26]. The former case has been widely used as a paradigmatic network example for the study of spreading dynamics taking place on the same

time-scale of network evolution [23, 24, 47, 48]. The models are described in detail as follows:

(i) Agent randomized (AR) null model

The AR null model only keeps the contact establishment mechanism that the recurrence intervals of the contacts for each individual are preserved. We keep all the human interaction time in the empirical datasets unchanged and replace each contact pair with two randomly selected individuals. Note that once a contact pair (u_1, v_1) is replaced by (u_2, v_2) , all the contacts between individuals u_1 and v_1 will be replaced by contact pair u_2 and v_2 .

(ii). Contact randomized (CR) null model

The CR null model only keeps the individual activity state mechanism that the degree and activity state of

individuals are unchanged in the network G_t at time step t . At each time step t , we randomly select two contacts associated with the four individuals and then rewire the two contacts. If the two contacts both have the same individual, we discard the contact pair and randomly select two new contacts. The rewiring steps are repeated more than $2L_C(t)$ times to ensure the rewiring of most contacts in G_t , where $L_C(t)$ is the number of contacts in G_t . After the rewiring procedure, the interaction time and recurrence interval of human contacts are completely different from those of original interaction datasets.

(iii) Activity-driven (AD) model

The AD model [22] is one of the most studied temporal network models. The generation of an AD model follows the rules. At each time step t , the G_t starts with N isolated nodes, and each node i is assigned an activity probability $a_i t_w$ to become active; then, the active nodes generate m links that are randomly connected to m other nodes, and the inactive nodes can only receive connections from active nodes; at next time step $t + 1$, all connections in G_t are omitted, and the steps are repeated. Here, we apply the AD model to generate temporal networks to match the real-world networks. The activity probability $a_i = \eta x_i$, where $x_i \in [\varepsilon, 1]$, ε is a lower cut-off that avoids possible divergences, and the x_i is drawn from a given probability function $F(x_i)$. The probability function $F(x_i)$ is statistically obtained from empirical data, and the average number of active agents per unit time is $\eta \langle x \rangle N$, where $\eta = 1$ in this work. The AD model does not capture the above two mechanisms in empirical data.

(iv) Activity driven with memory (ADM) model

A major shortcoming of the AD model lies in the total absence of correlations between the connections in consecutive time steps. The ADM model [26] extends the AD model by introducing a memory mechanism into to link creation. In the ADM model, each node remembers all its previously connected nodes. At each time step t , the G_t starts as in the AD model with N disconnected nodes, and each node i becomes active with probability $a_i t_w$. Each active node i interacts with one of the previously contacted n nodes with probability $p_{\text{ADM}} = n/(n + 1)$ and randomly connects a new node with probability $1 - p_{\text{ADM}}$.

In this work, the virus spreads on real human contact networks under a temporally periodic boundary condition (i.e., repeating the whole contact sequence) and network models until all the reachable agents are infected [17]. The five models have been introduced in detail in Sections 4.1 and 5.1. The relevant parameters of the MD model and AD model are measured in the corresponding datasets (see details on Tables 2 and 3). The parameters of the ADM

TABLE 3: The parameters of the activity-driven model.

| Activity-driven (AD) model | | | |
|----------------------------|------------|-----------|-----------|
| Parameters | Sex6yr | MIT_RM | School |
| $ \mathcal{N} $ | 16,706 | 96 | 236 |
| T | 64 | 230 | 104 |
| $F(x)$ | $Pl(-2.6)$ | $U(0, 1)$ | $U(0, 1)$ |
| ε | 10^{-2} | 10^{-2} | 10^{-2} |
| m | 1 | 5 | 3 |

$X \sim U(a, b)$ represents the uniform distribution from a to b .

model are the same as those of the AD model except the memory effect. We repeat the SI spreading process for all possible seeds and record the full prevalence time t_f , i.e., the time to reach 100% infection in a connected network or the largest connected component (LCC) of a disconnected network.

We calculate the ratio $\langle t_f^X(\lambda) \rangle / \langle t_f^{\text{real}}(\lambda) \rangle$ as a function of the infection rate λ , where $\langle t_f^{\text{real}} \rangle$ and $\langle t_f^X \rangle$ represent the average spreading time of a full prevalence in the real-world network *real* and the temporal network model X , respectively. If the ratio $\langle t_f^X(\lambda) \rangle / \langle t_f^{\text{real}}(\lambda) \rangle = 1$, the full prevalence time of a spreading in the real-world network is the same as that in the corresponding network model X . In other words, X is a reasonable temporal model for human contact networks in the study of spreading processes. As shown in Figure 8, the results in three empirical datasets are similar. The ratio $\langle t_f^{\text{AR}} \rangle / \langle t_f^{\text{real}} \rangle$ is always larger than 1 regardless of the infection rate λ , and the ratio $\langle t_f^{\text{AR}} \rangle / \langle t_f^{\text{real}} \rangle$ increases with the increase in the infection rate λ . The results illustrate that the individual state transition mechanism might promote the spreading processes since the spread of virus in the AR model, which only keeps the contact establishment mechanism and omits the individual activity state transition mechanism, is faster than that in real human contact networks. Moreover, the promotion of spreading is even larger when the infection rate λ is larger. However, there is a very small difference between the CR model and AD model, where the ratios $\langle t_f^{\text{CR}} \rangle / \langle t_f^{\text{real}} \rangle$ and $\langle t_f^{\text{AD}} \rangle / \langle t_f^{\text{real}} \rangle$ are both much smaller than 1. Both models are built without the contact establishment mechanism, which might cause a slowing-down effect on the spreading dynamics. Besides, the ratios $\langle t_f^{\text{AR}} \rangle / \langle t_f^{\text{real}} \rangle$ and $\langle t_f^{\text{ADM}} \rangle / \langle t_f^{\text{real}} \rangle$ are both larger than 1, provided that they both cannot capture the mechanism of individual activity state transition. We find that the ratio $\langle t_f^{\text{MD}} \rangle / \langle t_f^{\text{real}} \rangle$ always fluctuates around 1 and remains stable against the infection rate λ . Figure 8 confirms that the MD model is superior to all other models in characterizing human contact networks. This phenomenon highlights the crucial role of the two mechanisms considered in our model. They both serve as an indispensable component to characterize the evolution of the real-world networks. Binding them together enables us to accurately capture the contagion processes unfolding on empirical human contact networks.

5.2. *Dynamical Processes on Memory-Driven Networks.* We here study how the network evolution affects dynamic processes, such as the epidemic spreading on networks. The

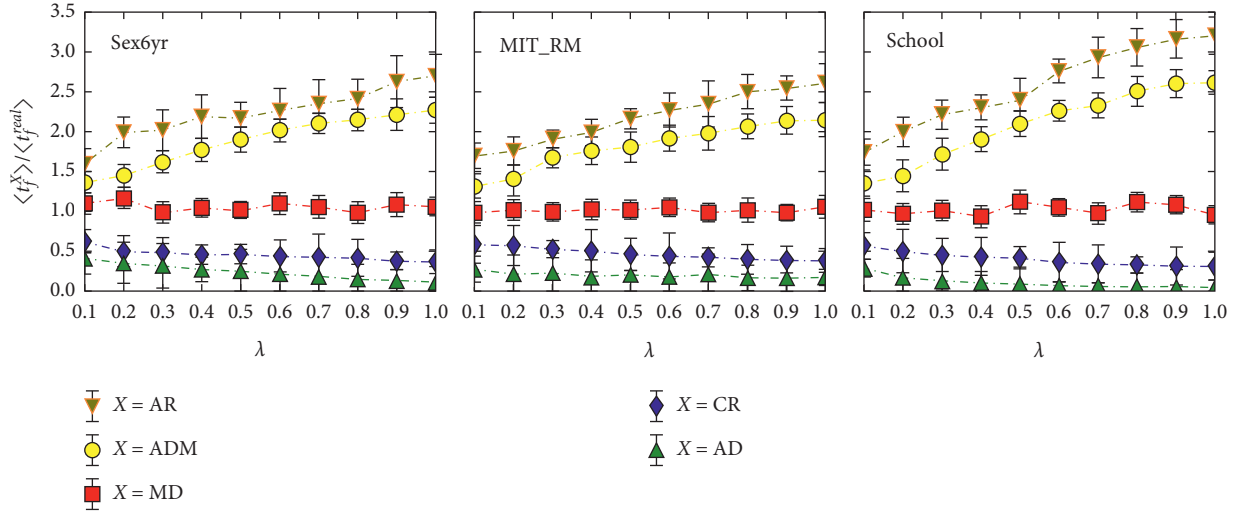


FIGURE 8: The ratio $\langle t_f^X \rangle / \langle t_f^{\text{real}} \rangle$ of the average full prevalence time as a function of the infection rate λ in three empirical datasets, where $X = \text{AR, CR, AD, ADM, MD}$ represents the agents randomized null models (∇), contacts randomized null models (\diamond), activity-driven models (\triangle), activity-driven models with memory (\hat{x}), and memory-driven models (\square), respectively. The simulations are performed on 10^3 realizations. The error bars here represent one binomial standard deviation from these averages with 95% confidence levels. All the models present the same critical spreading rate as the empirical counterparts.

exploring rate q in the contact establishment mechanism of the MD model can tune the evolution of network topology. We perform the SI spreading processes with infection rates $\lambda = 0.2, 0.4$, and 1.0 on temporal networks generated by the MD model with $N = 1000$ and the exploring rates $q = 0, 0.1, \dots, 1$, respectively, and record the full prevalence time t_f . Figure 9 indicates that t_f is extremely large when $q = 0$, which implies that all contacts in current time step are recurrences of previously established contacts; that is, no individual is willing to explore a new social relationship in the evolving network. Moreover, we observe that the average full prevalence time $\langle t_f \rangle$ declines rapidly with the exploring rate q , when q is small. When the exploring rate q further increases and reaches a specific value (around 0.1), the spreading time decreases slightly and finally meets a saturation. The fact illustrates that even when individuals establish new social ties with a small exploring rate q , the spreading process will significantly speed up; however, when q is larger than a specific value, the influence of q on the spreading process is rather limited.

6. Conclusion

We have presented the analysis of several empirical datasets of human contacts. We observe two crucial mechanisms, i.e., the individual activity state transition and contact establishment, which create the evolving structures of human contact networks. The individual activity state transition governs the activation of agents. The active agents establish contacts with others according to the strategies of contact establishment; that is, agents tend to interact with recently contacted one. We find that the transition probability $P_{\text{AI}}(k)$ or $P_{\text{AA}}(k)$ between the active state and the inactive state is a power-law function of the degree k . Moreover, the empirical datasets show that not all active nodes have the number of

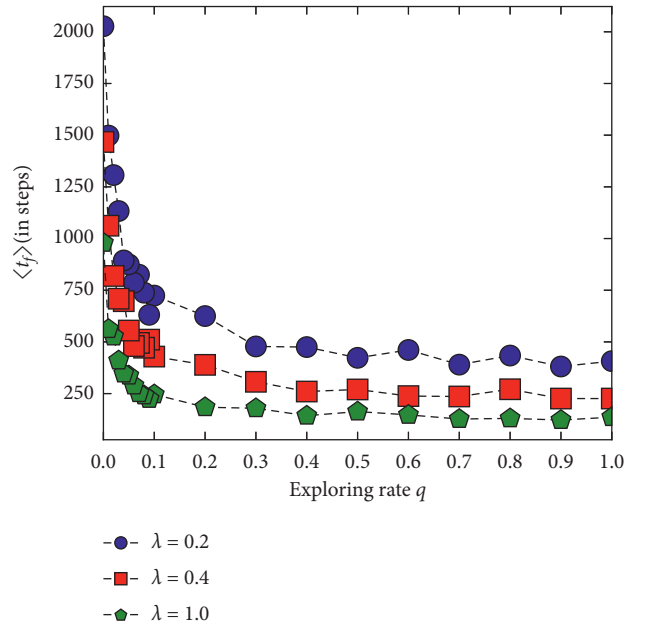


FIGURE 9: The average full prevalence time as a function of the exploring rate q versus the infection rate λ . The temporal networks are generated by the MD model with $N = 1000$, $A = 1$, $\alpha = 1$, $\gamma = 1.6$, $P_{\text{IA}} = 0.1$, $L = 40$, $T = 3000$, and s is sampled from $F(s) = (s/s_{\min})^{-\beta}$ with $\beta = 2.8$ and $s_{\min} = 1$. The simulation result is the average of 200 realizations.

contacts at each time step, and the recurrence intervals of individual contact establishment follow a power-law distribution. Considering the empirical observations, we propose a novel temporal network model (the memory-driven model) based on the two mechanisms. Furthermore, we study the effects of the two mechanisms on dynamical processes. We perform the SI spreading on the real-world

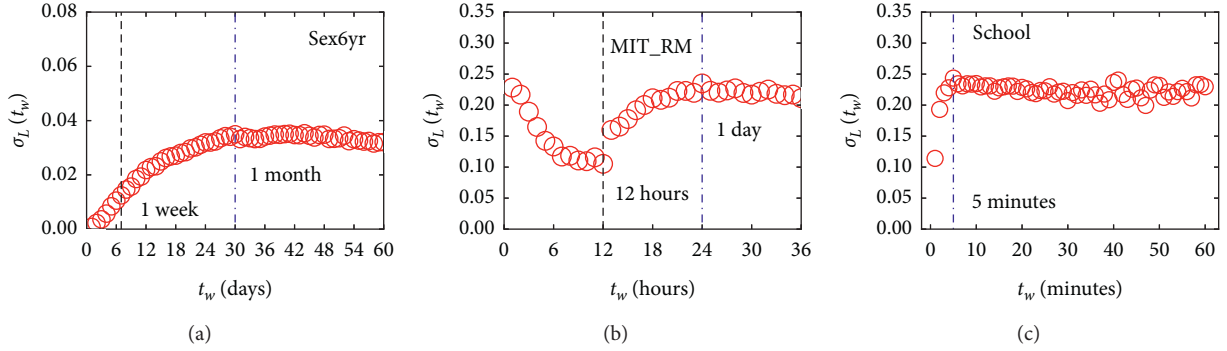


FIGURE 10: The average correlation as a function of the time window t_w . The values were averaged over all pairs of adjacent networks.

human contact networks and four temporal network models. The full prevalence times of SI spreading on the networks and models are compared. Our model shows a good agreement with three empirical temporal networks, which implies that the two mechanisms enable us to capture the evolution of human interactions in temporal networks. Interestingly, the results demonstrate that the individual activity state transition accelerates the diffusion processes; contrarily, the specific contact establishment strategy slows down the spreading. Besides, we find that the exploration of new social ties effectively promotes the spreading processes, and even a small exploring rate is sufficient to induce an explosive spreading on networks. The study thus paves the way to a better understanding of the mechanisms driving the evolution of human contacts and their effects on dynamic processes in real-world social systems.

Appendix

A. Data Description

- (i) The dataset of “Sex6yr” was gathered from an online forum in Brazil from September 2002 to October 2008. Note that although the data belong to the online survey, each record reflects a real offline sex trading activity among the sellers and buyers. We use the dataset after discarding the initial transient (312 days) of community growth [38]. The dataset is available in [39], and more details are described in [38].
- (ii) The dataset of “MIT_RM” was recorded by the Reality Mining Project conducted by the MIT Media Lab from September 2004 to May 2005. In the project, the subjects were required to use mobile phones with preinstalled software, which could sample their physical proximity via Bluetooth devices every 6 minutes and record the corresponding user tags. The dataset is available on the website of Reality Commons (<http://realitycommons.media.mit.edu>). From the dataset, we have removed 2 days as they are empty of human activities. We refer to [6] for more details on the data description and collection strategy.
- (iii) The dataset of “School” was collected in a French school on October 1, 2009, by the SocioPatterns

collaboration. The data recorded the time-resolved face-to-face proximity of children and teachers, with the Radio-Frequency IDentification (RFID) device embedded in badges. The dataset is available on the website of SocioPatterns (<http://www.sociopattern.org/datasets>), and more details are introduced in [40].

B. Selection of Time Window

The selection of time window size is crucial to analyse the evolution of the network structure. If t_w is too fine, the temporal contacts are aggregated over insufficient time. The resulting network is too sparse and messy, which makes it difficult to observe interesting phenomena such as the formation of a giant component or the disappearance of a cluster [13]. Conversely, if t_w is too coarse, the aggregated network will not be able to capture the critical temporal information such as link concurrency, time-respecting path, and reachability [49]. If t_w is large enough to aggregate all the contacts observed into a single time slice, the temporal network is degraded into a static network. Therefore, an appropriate time window should strike a balance between the distribution that disguise the relevant topological changes (small t_w) and the loss of temporal structural information (large t_w).

Inspired by the method introduced in [13, 50], here we consider the correlation between adjacent networks. First, we segment the empirical data into adjacency time steps of length $t_w = w\Delta t$, where Δt is the resolution and w is the number of resolutions in a time step. The contacts within the time interval $[(t-1)t_w, t \cdot t_w)$ are aggregated into a static undirected graph G_t . Once t_w is set, the temporal network is represented as a sequence of networks in time order $\mathcal{G}(t_w) = \{G_1, G_2, \dots, G_t, \dots, G_T\}$, where T is the total number of time steps. The network $G_t = (\mathcal{N}_t, \mathcal{L}_t)$ at time step t consists of a set of nodes \mathcal{N}_t connected by a set of links \mathcal{L}_t . Then, by the Jaccard index [13], we calculate the link overlap σ_L between adjacent networks $G_{t-1} = (\mathcal{N}_{t-1}, \mathcal{L}_{t-1})$ and $G_t = (\mathcal{N}_t, \mathcal{L}_t)$ as

$$\sigma_L(t, t_w) = \frac{|\mathcal{L}_{t-1} \cap \mathcal{L}_t|}{|\mathcal{L}_{t-1} \cup \mathcal{L}_t|}. \quad (\text{B.1})$$

σ_L takes values from the interval $[0, 1]$, with $\sigma_L = 0$ indicating that the adjacent networks share no common link and $\sigma_L = 1$ indicating that the same network is exactly retained, i.e., $G_t = G_{t-1}$. The average correlation across all adjacent networks can be calculated with

$$\sigma_L(t_w) = \frac{1}{T} \sum_{t=1}^T \sigma_L(t, t_w). \quad (\text{B.2})$$

As shown in Figures 10(a)–10(c), the average fraction of link overlap in each empirical dataset hits a peak when the time window t_w is selected as 1 month (Sex6yr), 1 day (MIT_RM), and 5 minutes (School), respectively. This phenomenon indicates that the adjacent networks are highly correlated under the corresponding time window, which characterizes the contact memory encoded in human interaction behaviors. As t_w further increases, the average fraction of link overlap begins to slowly decline. This can be explained by our observation that agents with a low activity gradually make random memoryless contacts and subsequently join into the aggregated network [29]. Therefore, according to the trade-off principle mentioned earlier, we aggregate the temporal contacts in empirical datasets with the time window corresponding to the peak point and denote it as w_0 . The time window size w_0 for three empirical datasets is selected as 1 month (Sex6yr), 1 day (MIT_RM), and 5 minutes (School), respectively.

Data Availability

(i) The dataset of “Sex6yr” was gathered from an online forum in Brazil from September 2002 to October 2008. Note that although the data belong to the online survey, each record reflects a real offline sex trading activity among the sellers and buyers. We use the dataset after discarding the initial transient (312 days) of community growth [38]. The dataset is available in [39], and more details are described in [38]. (ii) The dataset of “MIT RM” was recorded by the Reality Mining Project conducted by the MIT Media Lab from September 2004 to May 2005. In the project, the subjects were required to use mobile phones with preinstalled software, which could sample their physical proximity via Bluetooth devices every 6 minutes and record the corresponding user tags. The dataset is available on the website of Reality Commons (<http://realitycommons.media.mit.edu>). From the dataset, we have removed 2 days as they are empty of human activities. We refer to [6] for more details on the data description and collection strategy. (iii) The dataset of “School” was collected in a French school on October 1, 2009, by the SocioPatterns collaboration. The data recorded the time-resolved face-to-face proximity of children and teachers, with the Radio-Frequency IDentification (RFID) device embedded in badges. The dataset is available on the website of SocioPatterns (<http://www.sociopattern.org/datasets>), and more details are introduced in [40].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (nos. 71731004 and 61603097), in part by the Natural Science Fund for Distinguished Young Scholar of China (no. 61425019), and the Natural Science Foundation of Shanghai (no. 16ZR1446400).

References

- [1] M. S. Granovetter, “The Strength of Weak Ties,” *American journal of sociology*, vol. 78, pp. 1360–1380, 1973.
- [2] D. Krackhardt, N. Nohria, and B. Eccles, “Networks in the Knowledge Economy 82,” 2003.
- [3] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, “Modeling human dynamics of face-to-face interaction,” *Physical Review Letters*, vol. 110, Article ID 168701, 2013.
- [4] Y. Q. Zhang, X. Li, J. Xu, and A. Vasilakos, “Human interactive patterns in temporal network,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, pp. 214–222, 2015.
- [5] Y. Q. Zhang, X. Li, D. Liang, and J. Cui, “Transactions on technology and society,” *IEEE 2015 Index IEEE Communications Letters*, vol. 19, no. 12, pp. 1225–1228, 2013.
- [6] A. P. Nathan Eagle, “Reality mining: sensing complex social systems,” *Journal of Personal and Ubiquitous Computing*, vol. 12, pp. 255–268, 2006.
- [7] W. Dong, B. Lepri and A. S. Pentland, Modeling the co-evolution of behaviors and social relationships using mobile phone data,” in *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia (ACM)*, pp. 134–143, San Francisco, CA, USA, June 2011.
- [8] J.-P. Onnela, J. Saramäki, J. Hyvönen et al., “Analysis of a large-scale weighted network of one-to-one human communication,” *New Journal of Physics*, vol. 9, no. 6, p. 179, 2007.
- [9] L. Isella, J. Stehlé, A. Barrat et al., “Analysis of face-to-face behavioral,” *Journal of Theoretical Biology*, vol. 271, pp. 166–180, 2011.
- [10] H. H. Jo, M. Karsai, J. Kertész, and K. Kaski, “Circadian pattern and burstiness in mobile,” *New Journal of Physics*, vol. 14, Article ID 013055, 2012.
- [11] J. Saramäki and E. Moro, “From seconds to months: an overview of multi-scale,” *European Physical Journal B*, vol. 88, p. 164, 2015.
- [12] J. Saramäki, E. A. Leicht, E. López, S. G. Roberts, F. Reed-Tsochas, and R. I. Dunbar, “Persistence of social signatures in human communication,” *Proceedings of the National Academy of Sciences USA*, vol. 111, pp. 942–947, 2014.
- [13] V. Sekara, A. Stopczynski, and S. Lehmann, “Fundamental structures of dynamic social networks,” *Proceedings of the National Academy of Sciences USA*, vol. 113, pp. 9977–9982, 2016.
- [14] E. Ubaldi, N. Perra, M. Karsai, A. Vezzani, R. Burioni, and A. Vespignani, “Asymptotic theory of time-varying social networks,” *Scientific Reports*, vol. 6, 2016.
- [15] E. Ubaldi, A. Vezzani, M. Karsai, N. Perra, and R. Burioni, “Burstiness and tie activation strategies in time-varying,” *Scientific Reports*, vol. 7, 2017.
- [16] A. D. Medus and C. O. Dorso, “Memory effects induce structure in socia,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 9, Article ID p09009, 2014.
- [17] M. Karsai, M. Kivelä, R. K Pan et al., “Small but slow world: how network topology and burstiness,” *Physical Review E*, vol. 83, 2011.

- [18] G. Miritello, R. Lara, M. Cebrian, and E. Moro, "Analyzing dynamical social interactions as temporal networks," *Scientific Reports*, vol. 3, 2015.
- [19] C. L. Vestergaard, N. Voirin, A. Barrat et al., "Equilibrium states of generic quantum systems subject," *Physical Review E*, vol. 90, Article ID 42805, 2014.
- [20] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, and F. Schweitzer, "Causality-driven slow-down and speed-up of diffusion," *Nature Communications*, vol. 5, p. 5024, 2014.
- [21] L. E. Rocha and V. D. Blondel, "Bursts of vertex activation and epidemics in evolving," *PLoS Computational Biology*, vol. 9, Article ID e1002974, 2013.
- [22] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, "Activity driven modeling of time varying networks," *Scientific Report*, vol. 2, p. 469, 2012.
- [23] K. Hoppe and G. Rodgers, "Mutual selection in time-varying networks," *Physical Review E*, vol. 88, 2013.
- [24] I. Pozzana, K. Sun, and N. Perra, "Epidemic spreading on activity-driven networks," *Physical Review E*, vol. 96, 2017.
- [25] L. Alessandretti, K. Sun, A. Baronchelli, and N. Perra, "Random walks on activity-driven networks with attractiveness," *Physical Review E*, vol. 95, 2017.
- [26] M. Karsai, N. Perra, and A. Vespignani, "Time varying networks and the weakness of strong ties," *Scientific Report*, vol. 4, 2014.
- [27] H. Kim, M. Ha, and H. Jeong, "Scaling properties in time-varying networks with memory," *The European Physical Journal B*, vol. 88, pp. 1–8, 2015.
- [28] E. Valdano, C. Poletto, A. Giovannini, D. Palma, L. Savini, and V. Colizza, "Predicting epidemic risk from past temporal contact," *PLoS Computational Biology*, vol. 11, Article ID e1004152, 2015.
- [29] J. Li, C. Li, and X. Li, "Quantifying the contact memory in temporal human interactions," in *Proceedings of the 2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, Florence, Italy, May 2018.
- [30] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.
- [31] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, p. 3200, 2001.
- [32] C. Castellano and R. Pastorsatorras, "Thresholds for epidemic spreading in networks," *Physical Review Letters*, vol. 105, Article ID 218701, 2010.
- [33] J. B. Wang, L. Wang, and X. Li, "Identifying Spatial Invasion of Pandemics on Metapopulation," *IEEE Transactions on Cybernetics*, vol. 46, pp. 2782–2795, 2016.
- [34] B. Qu, C. Li, P. V. Mieghem, and H. Wang, "Ranking of nodal infection probability in susceptible," *Scientific Report*, vol. 7, p. 9233, 2017.
- [35] J. L. Iribarren and E. Moro, "Impact of human activity patterns on the dynamics," *Physical Review Letters*, vol. 103, Article ID 038702, 2009.
- [36] A. Montanari and A. Saberi, "The spread of innovations in social networks," *Proceedings of the National Academy of Sciences USA*, vol. 107, 2010.
- [37] G. E. Kreindler and H. P. Young, "Rapid innovation diffusion in social networks," *Proceedings of the National Academy of Sciences USA*, vol. 111, pp. 10881–10888, 2014.
- [38] L. E. C. Rocha, F. Liljeros, and P. Holme, "Small molecules of different origins have distinct," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 5706–5711, 2007.
- [39] L. E. Rocha, F. Liljeros, and P. Holme, "Simulated epidemics in an empirical spatiotemporal," *PLoS Computational Biology*, vol. 7, Article ID e1001109, 2011.
- [40] J. Stehlé, N. Voirin, A. Barrat et al., "Structural insights into the mechanism of protein," *PLoS One*, vol. 6, Article ID 23176, 2011.
- [41] A. L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, pp. 207–211, 2005.
- [42] A. Moinet, M. Starnini, and R. Pastor-Satorras, "Burstiness and aging in social temporal networks," *Physical Review Letters*, vol. 114, 2015.
- [43] P. Erdos and A. Rényi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–60, 1960.
- [44] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London A: Mathematical, Physical And Engineering Sciences*, vol. 121, pp. 700–721, 1927.
- [45] R. M. Anderson, R. M. May, and B. Anderson, *Infectious Diseases of Humans: Dynamics and Control*, Wiley Online Library, Hoboken, NJ, USA, 1992.
- [46] M. Starnini, A. Machens, C. Cattuto, A. Barrat, and R. Pastor-Satorras, "Evolving protein interaction networks from gene duplication," *Journal of Theoretical Biology*, vol. 337, pp. 89–100, 2013.
- [47] A. Moinet, R. Pastor-Satorras, and A. Barrat, "Effect of risk perception on epidemic spreading in temporal," *2018 Physical Review E*, vol. 97, Article ID 012313, 2018.
- [48] Y. Zhang, J. Wang, C. Li, and X. Li, "Epidemic spreading in time-varying networks with activity-driven infectivity," in *Proceedings of the 2019 IEEE 58th Conference on Decision and Control (CDC)*, Nice, France, December 2019.
- [49] P. Holme and J. Saramäki, "Temporal networks," *Physics Reports*, vol. 519, no. 3, pp. 97–125, 2011.
- [50] R. Sulo, T. Berger-Wolf, and R. Grossman, "Meaningful selection of temporal resolution for dynamic networks," in *Proceedings of the Eighth Workshop on Mining and Learning with Graphs (ACM)*, pp. 127–136, Washington DC, USA, March 2010.

Research Article

Policy and Law Assessment of COVID-19 Based on Smooth Transition Autoregressive Model

Jieqi Lei,¹ Xuyuan Wang,² Yiming Zhang,³ Lian Zhu ,⁴ and Lin Zhang ²

¹School of Humanities, Beijing University of Posts and Telecommunications, Beijing 100876, China

²School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

³International School, Beijing University of Posts and Telecommunications, Beijing 1000876, China

⁴School of Journalism and Communication, Shanghai International Studies University, Shanghai 200083, China

Correspondence should be addressed to Lian Zhu; zhulian@shisu.edu.cn and Lin Zhang; zhanglin2011@bupt.edu.cn

Received 16 October 2020; Revised 4 December 2020; Accepted 22 December 2020; Published 20 January 2021

Academic Editor: Xiaoke Xu

Copyright © 2021 Jieqi Lei 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.

As of the end of October 2020, the cumulative number of confirmed cases of COVID-19 has exceeded 45 million and the cumulative number of deaths has exceeded 1.1 million all over the world. Faced with the fatal pandemic, countries around the world have taken various prevention and control measures. One of the important issues in epidemic prevention and control is the assessment of the prevention and control effectiveness. Changes in the time series of daily new confirmed cases can reflect the impact of policies in certain regions. In this paper, a smooth transition autoregressive (STAR) model is applied to investigate the intrinsic changes during the epidemic in certain countries and regions. In order to quantitatively evaluate the influence of the epidemic control measures, the sequence is fitted to the STAR model; then, comparisons between the dates of transition points and those of releasing certain policies are applied. Our model well fits the data. Moreover, the nonlinear smooth function within the STAR model reveals that the implementation of prevention and control policies is effective in some regions with different speeds. However, the ineffectiveness is also revealed and the threat of a second wave had already emerged.

1. Introduction

In 2020, the COVID-19 epidemic is changing the way how people live, work, study, and socialize [1, 2]. In the absence of effective vaccines, only nonpharmaceutical public health interventions can be used to break the chain of transmission of the virus [3, 4]. Huge social and economic cost measures are conducted, such as the “lockdown” of severe regions, the establishment of adequate isolation places, the closure of nonessential public venues, and the closure of schools and workplaces [5–10].

China has adopted a series of strict control measures, including the lockdown of Wuhan starting from January 23, 2020, the establishment of Vulcan Mountain Hospital, Thor Mountain Hospital, and mobile cabin hospital, calling for the public to stay at home, not dining together, not visiting friends and relatives, working from home for adults, studying online at home for kids and college students, and

tracing and isolating close contacts. The virus spread alarmingly fast in late January in China. Based on the public reports on the number of confirmed cases, the prevalence of COVID-19 outside Hubei Province came to a controllable size in late February [11–14]. The epidemic has been under effective control by March 2020. The Wuhan lockdown eventually ended on April 8, 76 days since its commencement. Moreover, China has timely contained several scattered confirmed cases in Jilin, Beijing, Xinjiang, Liaoning, and other places afterwards.

During the pandemic, all countries have implemented various prevention and control policies. Faced with a novel coronavirus that is highly contagious without reliable vaccine, the effectiveness of policies in different countries and regions is an important and urgent topic. Gregory et al. [15] provided evidence that state-wide mandates are effective in promoting social distancing in America through Google big data. Saez et al. [16] used a time-series design and a

generalized linear hybrid model to explain and estimate the changes in the epidemic curve following quarantine measures in Spain. Nicholas et al. [17] used a random age structure propagation model to explore the effects of a range of interventions in UK and found that all four basic interventions were likely to reduce the basic reproductive number. It has been shown that in most countries the number of daily confirmed cases and the number of daily deaths show signs of decline from 1 to 4 weeks after the highest levels of social distancing measures were enacted and that the effect of social distancing measures on COVID-19 transmission varies among countries [18].

The effectiveness of prevention and control measures will be directly reflected in the trend of the epidemic. In order to quantitatively identify the inflection points hidden in the time series of daily new confirmed cases and further evaluate the effectiveness, a smooth transition autoregressive (STAR) model is utilized to analyze the epidemic data in different countries and regions.

Transitions are usually characterized by nonlinear models [19–21]. There are three main types of nonlinear models: the ARCH model proposed by Engle [22], the Markov mechanism transition model proposed by Hamilton with discrete variations of dependent variables determined by Markov chain [23], and the smooth transition autoregressive model proposed by Granger and Terasvirta [24], namely, the STAR model, while Tong's autoregressive model of the gate can be regarded as a special case of STAR [25]. ARCH model is mainly to describe the volatility of nonlinear. The Markovian transition model needs much information to infer the state of the variables, and the result can only infer the probability of the transformation zone, without specifically describing the nonlinear form of the transformation. The STAR model can realize the continuous smooth nonlinear transition and gives the nonlinear form of the mechanism transformation. Compared with other models, our method can not only well fit the epidemic curve but also quantify the inflection point, as well as the speed of transition, of the epidemic time series without external or prior information.

Combining with the release time of policies, we can provide the evaluation of the effectiveness of epidemic prevention and control measures. In fact, the inflection points are related to the quantitative parameters reflecting the switch of expansion and contraction within the sequences. It is revealed by the nonlinear function in the STAR model that the effectiveness of certain policies usually showed up within 2 weeks to 2 months. Ineffective policies have little impact on the trend of time series within 2 months or more.

An outline of this paper is as follows. The data and methodology are given in Section 2. Results are illustrated in Section 3. Discussions are given in Section 4.

2. Data and Methodology

The daily confirmed cases of COVID-19 in different regions, which are Guangdong Province in China, Brazil, France, Germany, India, Italy, Mexico, Spain, and the United States

of America, are collected from the World Health Organization database. Before fitting to our model, the original time series are smoothed in order to reduce the fluctuation and keep the trend. In addition, measures and policies to prevent the transmission of COVID-19 in different regions are collected from national or regional official websites and news websites. The information and time series of daily confirmed cases are combined together to analyze the relationship between the social measures and the intrinsic changes of daily confirmed cases.

The daily new confirmed COVID-19 sequence data are modeled by smooth transition autoregressive model (STAR). The STAR model is divided into the autoregressive part and the nonlinear smooth part; the specific form of the model is as follows:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + (\beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p}) G(\gamma, c; y_{t-d}) + \varepsilon_t, \quad (1)$$

where $(\alpha_0, \alpha_1, \dots, \alpha_p)$ and $(\beta_0, \beta_1, \dots, \beta_p)$ are vector parameters, and y_t is the response variable, $(y_{t-1}, \dots, y_{t-p})$ is the vector of the explanatory variables. $G(\gamma, c; y_{t-d}) = 1 - \exp(-\gamma(y_{t-d} - c)^2)$, where $\gamma > 0$ is the transition speed, and it is the function $G(\gamma, c; y_{t-d})$ that causes the nonlinearity of the model. It is continuous with the observable variable y_{t-d} . As y_{t-d} changes, $G(\gamma, c; y_{t-d})$ changes between 0 and 1 smoothly. The parameter $\gamma > 0$ determines the speed of the smooth transition. Moreover, parameter c measures the position where the smooth transition takes place within the system under consideration, which can be viewed as the threshold for different growing patterns. Finally, p is the order of time lag for autoregression, and d is the time lag where state variable y_{t-d} is converted.

In the application of the STAR model, all the parameters p , d , γ , c , α_i , and β_i , $i = 0, 1, \dots, p$, should be estimated, in which the values of p and d should be determined first, and then the remaining parameters are estimated. In this work, the time lag parameters are set as $p = 7$ or $p = 8$, and $d = 4$ or $d = 5$ for most cases due to the weekly stability of the time series and the statistical feature of incubation period.

The STAR model gives good fitting of the time series in consideration, and the associated nonlinear function G tells the transition within the sequence. When a system is at a lower level, it is easy to cause an expansion, which makes the system expand; thus, $G(\gamma, c; y_{t-d})$ would decrease. On the other hand, when the system is at a higher level, it tends to decline; hence, the value of $G(\gamma, c; y_{t-d})$ would increase accordingly. The larger the value of $G(\gamma, c; y_{t-d})$ was, the more likely transition of expansion or contraction would take place. Therefore, the intrinsic change within the sequence of the daily confirmed cases can be obtained from the value of $G(\gamma, c; y_{t-d})$. The reference threshold for G is usually 0.5.

In order to evaluate the effect of the prevention measures, the comparison between the release time of a policy and the time that transition takes place in function G can accomplish our task. When function G goes across the threshold, intrinsic transition happens, and hence the

effectiveness of policies is proven as long as the change in G happens within a certain time span after the release of policies. Finally, the quantification of the error in our fitting result is the mean absolute value defined as

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|, \quad (2)$$

where $\{y_t, t = 1, \dots, n\}$ is the time series of the real data and $\{\hat{y}_t, t = 1, \dots, n\}$ is the sequence of the fitted values from our model.

3. Results

According to the STAR model, the data of daily new cases in each region are fitted to the model. Moreover, the nonlinear function is obtained simultaneously to quantify the transition points within the sequence. The impact of policies on the epidemic will be analyzed by combining the transitions in the time series with the timeline of epidemic prevention policies released and implemented in different regions. Guangdong Province in China is analyzed first to show the validation of our STAR model and the positive impact of epidemic prevention policies in detail. Then, cases in different countries are analyzed afterwards. We focused on the effectiveness as well as the ineffectiveness conducted by typical countries. Moreover, we take a look at the second wave of COVID-19 in two European countries with our STAR model, resulting in calling for closer attention and more effective and stricter prevention policies.

3.1. Effective Policies

3.1.1. Guangdong Province of China. The Chinese government adopted a series of epidemic prevention policies decidedly, which have successfully contained the epidemic and set a successful experience. We selected the data of daily new confirmed cases from January 24 to March 2 in Guangdong Province, China. The moving average is set as 5 to smooth the fluctuation. The lag order for autoregression is set as $p = 3$ due to the short epidemic period and the lag order for the nonlinear transition is set as $d = 1$ due to the fast reaction. Then, the nonlinear fitting equation can be obtained as follows:

$$\begin{aligned} \hat{y}_t = & 76.6 - 4.2y_{t-1} + 7.6y_{t-2} - 3.8y_{t-3} \\ & + (-76.8 + 5.7y_{t-1} - 8.0y_{t-2} + 3.6y_{t-3}) \\ & \cdot [1 - \exp(-24.0(y_{t-1} - 50.8)^2)]. \end{aligned} \quad (3)$$

Figure 1 shows the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). Given by the STAR model, the transitions within the sequence of daily new cases are obtained from function G .

To be specific, let us see the fitting results in more detail. The transition speed parameter $\gamma = 24.0$ measures the speed of the structural transition, which indicates that the epidemic situation in Guangdong Province was characterized

by rapid reaction. The duration of the transition within function G is 2 weeks, which is the common time length for quarantine of susceptible people. $c = 50.8$ indicates that about 50.8 new cases daily is the switch boundary of expansion and contraction. In fact, Guangdong had quickly adopted a series of antiepidemic measures during the period from January 24 to February 4. The dates that policies released are represented in Figure 1 with vertical dashed lines. According to the STAR model, the changes in G reflect the intrinsic changes within the time series.

The control policies are effective by comparing the changes in G and the implementation of the policies shown in Figure 1. The effective policies in Guangdong Province represent cases in most Chinese provinces, showing that the epidemic can be quickly contained as long as epidemic prevention policies are timely and strictly conducted. The model fitting error is shown in Figure 2. The mean absolute error (MAE) for Guangdong is 2.3, which shows the good fitting of the STAR model. In the following, typical countries with effective containment policies are analyzed.

3.1.2. Italy. The daily new cases from February 22 to May 30, 2020, were selected for Italy. The moving average order is set as 7 to smooth weekly fluctuation. The lag orders for autoregression and nonlinear function are set as $p = 7$ and $d = 5$, respectively. The nonlinear fitting equation is as follows:

$$\begin{aligned} \hat{y}_t = & 92.4 + 1.2y_{t-1} - 0.1y_{t-2} - 0.2y_{t-3} \\ & + 0.4y_{t-5} - 0.4y_{t-6} + 0.1y_{t-7} \\ & + (-62.9 - 0.2y_{t-1} + 0.3y_{t-2} \\ & + 0.1y_{t-4} - 0.1y_{t-5} + 0.2y_{t-6} + 0.5y_{t-7}) \\ & \cdot [1 - \exp(-5.3(y_{t-5} - 3261.2)^2)]. \end{aligned} \quad (4)$$

The curves for the real data, fitting results, and G function are shown in Figure 3. The smooth transition in G takes place near $c = 3261.2$, and the speed parameter is $\gamma = 5.3$. The mean absolute error for Italy is 38.7 (Figure 4).

The impact of policies in Italy is positive. Compared to the fitting result of Guangdong Province with $\gamma = 24.0$ and $c = 50.8$, the speed for Italy is $\gamma = 5.3$ and $c = 3261.2$. Due to the large amount of daily confirmed cases, the boundary of transition is around 3261.2. It is difficult to prevent the pandemic when the scale is large. Therefore, the speed is slower. Since March 8, Italy has adopted extensive lockdown measures. Specific measures to restrict contacts were first implemented in the northern region and then conducted nationwide on March 10. As illustrated in Figure 3, the duration between two transitions indicated by function G is nearly 1 month to get the daily new confirmed cases in Italy changing to shrinking period.

3.1.3. Germany. The data of daily new confirmed cases from February 28 to May 30, 2020, were selected for Germany. The moving average order is set as 7 to smooth the weekly periodic fluctuation. The lag orders for autoregression and nonlinear transition are set as $p = 8$

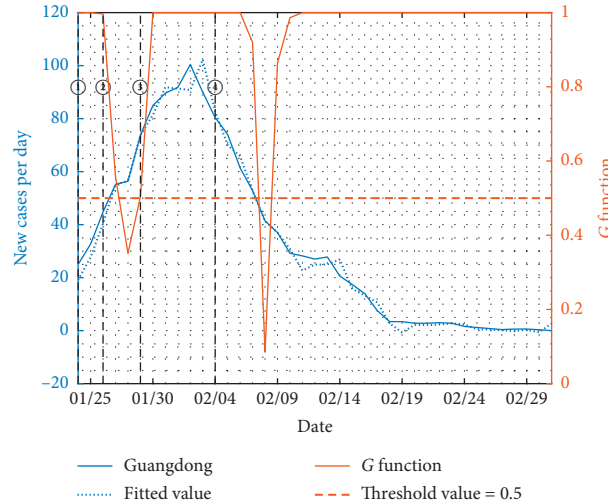


FIGURE 1: The fitting result of Guangdong Province, China. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① close entertainment and gatherings, ② close other places, ③ isolate confirmed cases, and ④ suspend urban traffic.

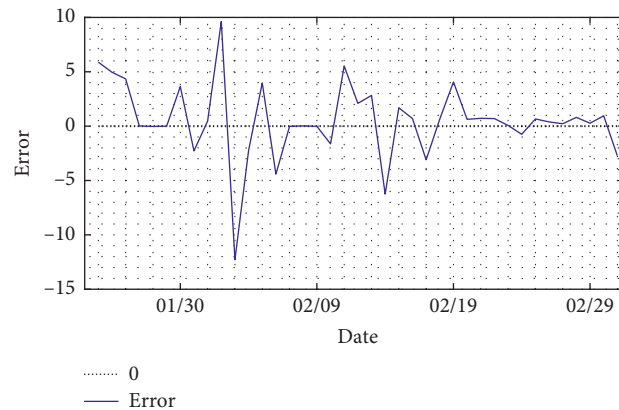


FIGURE 2: The fitting error of Guangdong Province, China. The mean absolute error is 2.3, indicating the goodness of fit.

and $d = 4$, respectively. Then, we can get the nonlinear fitting equation as follows:

$$\begin{aligned} \hat{y}_t = & -314.3 + 0.8y_{t-1} + 0.6y_{t-2} - 0.5y_{t-3} + 0.2y_{t-4} - 0.1y_{t-5} \\ & + 0.5y_{t-6} - 0.8y_{t-7} + 0.3y_{t-8} \\ & + (358.0 + 0.5y_{t-1} - 0.9y_{t-2} + 0.7y_{t-3} \\ & - 0.1y_{t-4} - 0.3y_{t-5} - 0.5y_{t-6} + 0.4y_{t-7} + 1.7y_{t-8}) \\ & \cdot [1 - \exp(-2.2(y_{t-4} - 2952.8)^2)]. \end{aligned} \quad (5)$$

The model fitting results and errors are shown in Figures 5 and 6. The location of smooth transition is near $c = 2952.8$, and the transition speed parameter is $\gamma = 2.2$. The mean absolute error is 42.

The transition speed parameter $\gamma = 2.2$ shows that the containment in Germany is slow. Moreover, the transition boundary $c = 2952.8$ indicates a severe situation. On March

8, 2020, the number of total confirmed cases in Germany exceeded 1,000. A week later, schools were closed. The government raised the risk level to “high.” When the number of total confirmed cases exceeded 30,000 on March 23, the German government tightened social controls on activities and gatherings. The implementation of these policies made the daily confirmed cases decline and resulted in changes in G . The duration between the first and last transitions in G is nearly 40 days, indicating the effectiveness of prevention and hard work for Germany to get the pandemic under control.

3.1.4. France. The data of daily new confirmed cases between February 26 and May 28, 2020, were selected for France, and the order of moving average is set as 7 to smooth the weekly periodic fluctuation. The lag orders for autoregression and nonlinear transition are set as $p = 7$ and $d = 4$, respectively. The nonlinear fitting equation is as follows:

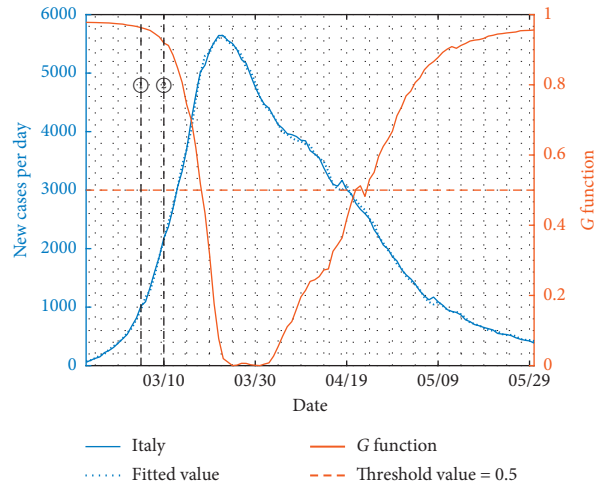


FIGURE 3: The fitting result of Italy. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① close schools across the country and ② nationwide lockdown.

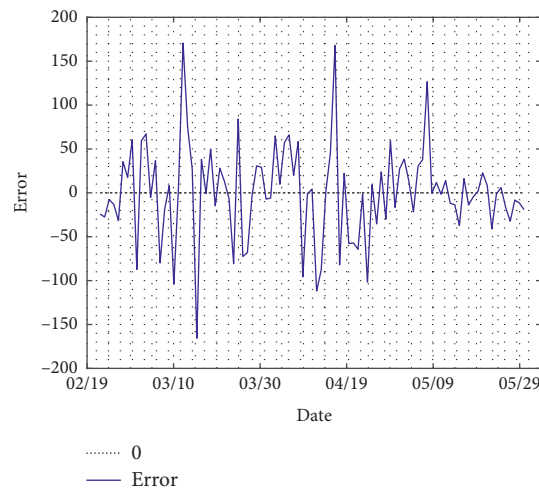


FIGURE 4: The fitting error of Italy with a mean absolute error of 38.7.

$$\begin{aligned} \hat{y}_t = & 6728.3 - 1.7y_{t-1} - 0.7y_{t-2} - 0.4y_{t-3} + 1.0y_{t-4} \\ & + 1.4y_{t-5} + 4.5y_{t-6} - 3.4y_{t-7} + (-6688.4 + 2.7y_{t-1} + 0.7y_{t-2} + 0.5y_{t-3} - 1.0y_{t-4} - 1.4y_{t-5} - 4.6y_{t-6} + 3.3y_{t-7}) \\ & \cdot [1 - \exp(-13.8(y_{t-4} - 2890.3)^2)]. \end{aligned} \quad (6)$$

The fitting results and errors are shown in Figures 7 and 8. The location of smooth transition is near $c = 2890.3$, and the transition speed parameter is $\gamma = 13.8$. The mean absolute error is 96. The critical value for transition is $c = 2890.3$, which is similar to that of Italy and Germany. The speed parameter $\gamma = 13.8$, indicating a fast effectiveness of a series of policies released by the French government. In fact, the duration between the first and last transitions of G is

around 20 days, as shown in Figure 7, showing the fast and effective prevention measures in France.

3.1.5. Spain. The data of daily new confirmed cases for Spain were selected from February 25 to May 30, 2020. The moving average is set as 7 to smooth the weekly fluctuation. The lag orders for autoregression and nonlinear transition are set as

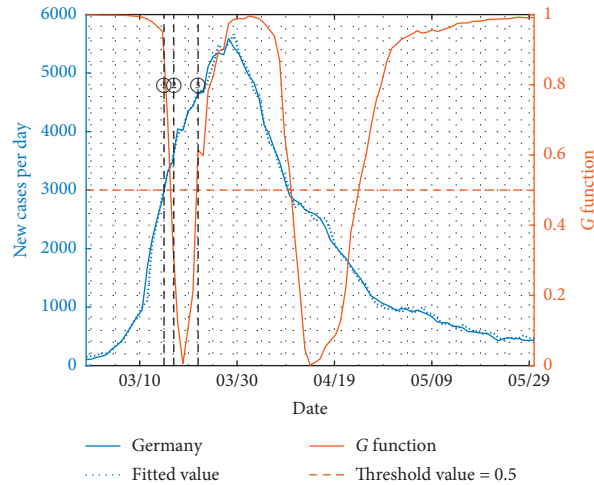


FIGURE 5: The fitting result of Germany. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① close schools across the country, ② upgrade the risk level of the outbreak to “high,” and ③ tighten controls on social activities and gatherings.

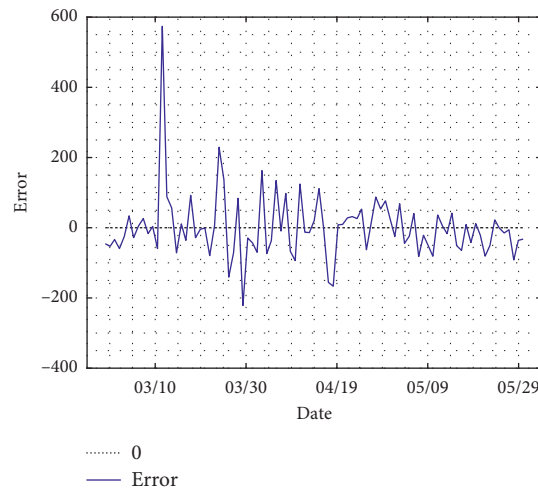


FIGURE 6: The fitting error of Germany with a mean absolute error of 42.

$p = 7$ and $d = 4$. Then, the nonlinear fitting equation is as follows:

$$\begin{aligned} \hat{y}_t = & 526.9 + 0.9y_{t-1} - 0.5y_{t-2} + 1.2y_{t-3} - 0.8y_{t-4} + 1.6y_{t-5} - 1.3y_{t-6} - 0.2y_{t-7} \\ & + (-494.4 + 0.6y_{t-1} + 0.3y_{t-2} - 1.4y_{t-3} + 1.1y_{t-4} - 2.1y_{t-5} + 1.5y_{t-6} + 0.2y_{t-7}) [1 - \exp(-4.9(y_{t-4} - 3618.1)^2)]. \end{aligned} \tag{7}$$

The model fitting results and errors are shown in Figures 9 and 10. The transition speed parameter is $\gamma = 4.9$, indicating a slow transition and prevention effect. The boundary for smooth transition in Spain is around $c = 3618.1$, showing a severe situation to combat the

pandemic. The mean absolute error is 85, showing a good fitting result.

Spain imposed a regional lockdown on March 16. Two weeks later, on March 30, Spain imposed a stricter restriction aiming at reducing mobility and nonessential industrial

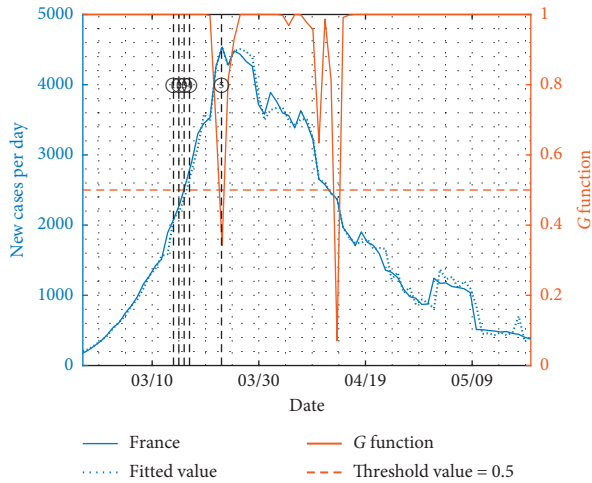


FIGURE 7: The fitting result of France. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① close all nonessential areas, ② declare the highest stage of epidemic prevention and close all nonessential public places, ③ close all schools, ④ close borders, and ⑤ put field hospitals into operation.

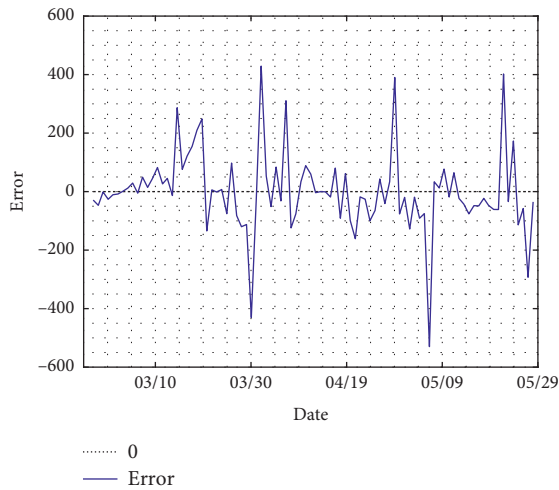


FIGURE 8: The fitting error of France with a mean absolute error of 96.

activity throughout the country. However, it took nearly two months for Spain to overcome the first wave of pandemic, resulting from the duration between the first and the last transitions of G . The transition speeds for Spain and Italy are 4.9 and 5.3, which are similar. However, it took nearly twice the time period for Spain to get the pandemic under control as that of Italy. The essence for the difference lies in the hesitation towards imposing strict restriction at early stage of the epidemic. The epidemic between March 16 and March

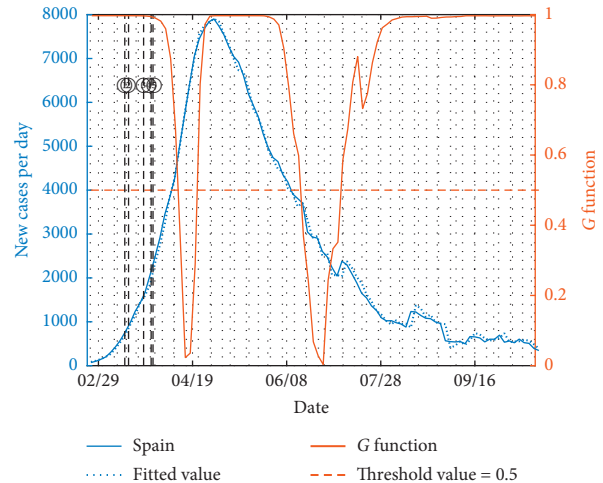


FIGURE 9: The fitting result of Spain. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① restrict free traveling for 15 days, ② isolate residents at home and close nonessential places, ③ transform Madrid’s convention and exhibition center into a “makeshift hospital,” ④ force workers in noncore industries to stay at home, and ⑤ upgrade epidemic prevention and control measures.

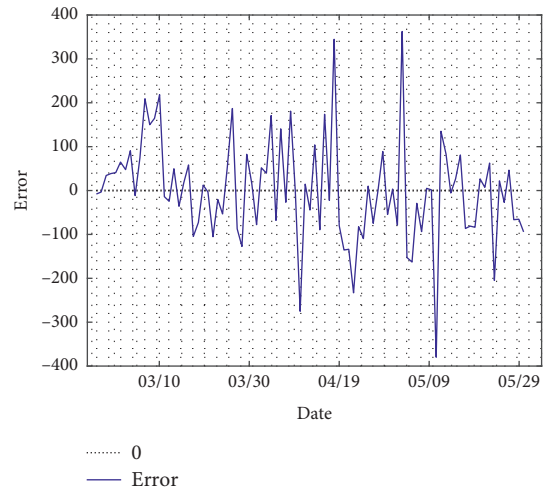


FIGURE 10: The fitting error of Spain with a mean absolute error of 85.

30 showed the ineffectiveness of the early control measures.

3.1.6. *United States of America.* The data of daily new confirmed cases for USA were selected from March 6 to May 24, 2020. The moving average order is set as 7 to smooth the weekly fluctuation. The lag orders for autoregression and nonlinear transition are set as $p = 7$ and $d = 4$, respectively. The nonlinear fitting equation is as follows:

$$\begin{aligned} \hat{y}_t = & 34733.1 - 1.1y_{t-1} - 4.8y_{t-2} + 4.8y_{t-3} - 3.7y_{t-4} - 0.2y_{t-5} - 4.1y_{t-6} + 4.4y_{t-7} \\ & + (-34058.6 + 2.0y_{t-1} + 4.8y_{t-2} - 4.9y_{t-3} + 4.8y_{t-4} - 0.3y_{t-5} - 4.3y_{t-6} - 4.0y_{t-7}) \left[1 - \exp(-4.5(y_{t-4} - 15942.5)^2) \right]. \end{aligned} \quad (8)$$

The model fitting results and errors are shown in Figures 11 and 12. The transition speed parameter is $\gamma = 4.8$, indicating a slow transition. The critical value for transition is around $c = 15942.5$, reflecting a severe situation in USA. The mean absolute error for the fitting is 431.

The nonlinear transition function G in Figure 11 showed only one transition. The intrinsic contraction had not arrived till May 24, 2020. On April 17, the reopening policies were released aiming at recovering economy in three stages. There is no significant fluctuation short after the reopen. This indicates the possibility of reopening under close prevention. However, it is still at high risk of further outbreak since the system had not switched to contraction stage. As we know, after the death of Freud, the trend grew

$$\begin{aligned} \hat{y}_t = & 263.1 + 1.3y_{t-1} - 0.5y_{t-2} - 0.1y_{t-3} + 0.3y_{t-5} + 0.3y_{t-6} - 1.3y_{t-7} \\ & + (-265.1 - 22.6y_{t-1} + 18.8y_{t-2} + 30.5y_{t-3} + 12.4y_{t-4} - 18.8y_{t-5} - 20.0y_{t-6} + 43.5y_{t-7}) \left[1 - \exp(-4.8(y_{t-4} - 27722.5)^2) \right]. \end{aligned} \quad (9)$$

The model fitting results and errors are shown in Figures 13 and 14. The transition speed parameter is $\gamma = 4.8$, indicating a slow transition. Due to the severe pandemic, the boundary between expansion and contraction is around $c = 27722.5$. The mean absolute error is 272.

In order to prove the effectiveness of certain policies, there must be changes within a period containing as short incubation period as possible. Since the common quarantine time is 14 days, the change should show up within 1 month or so. However, for the case of Brazil, not only the trend of the time series but also the G function remained unaffected

$$\begin{aligned} \hat{y}_t = & -23643.3 + 0.8y_{t-1} + 0.2y_{t-2} - 0.4y_{t-3} - 0.4y_{t-4} - 0.3y_{t-5} + 1.8y_{t-6} - 0.2y_{t-7} \\ & + (23681.1 + 0.7y_{t-1} - 0.9y_{t-2} + 1.0y_{t-3} - 0.2y_{t-4} + 0.8y_{t-5} - 2.4y_{t-6} + 0.4y_{t-7}) \left[1 - \exp(-25.8(y_{t-4} - 59008.5)^2) \right]. \end{aligned} \quad (10)$$

The model fitting results and errors are shown in Figures 15 and 16. The transition speed parameter is $\gamma = 25.8$, indicating a quick transition. The critical value between expansion and contraction states is $c = 59008.5$. The mean absolute error is 154 for our fitting.

explosively due to demonstrations and gatherings. Close contacts made the transmission of coronavirus easy and resulted in further spread of the pandemic.

3.2. Ineffective Policies

3.2.1. Brazil. The data of daily new confirmed cases from March 5 to June 30, 2020, were selected for Brazil. The moving average order is set as 7 to smooth the weekly fluctuation. The lag orders of autoregression and nonlinear transition are set as $p = 7$ and $d = 4$, respectively. The nonlinear fitting equation is as follows:

for more than 2 months after the release of policies. The STAR model shows a good fitting, as well as the ineffectiveness of the prevention measures.

3.2.2. India. The data of daily new confirmed cases were selected from March 3 to September 20, 2020, for India. The moving average is set as 7 to smooth the weekly periodic fluctuation. The lag orders of autoregression and nonlinear transition are set as $p = 7$ and $d = 4$, respectively. The nonlinear fitting equation is as follows:

The situation of India is similar to that of Brazil. However, it possesses longer time that the sequence trend and the transition function G remained unaffected, which was more than 3 months, refer to Figure 15. The transition in G indicated the slowdown of increase trend caused by potential natural growth.

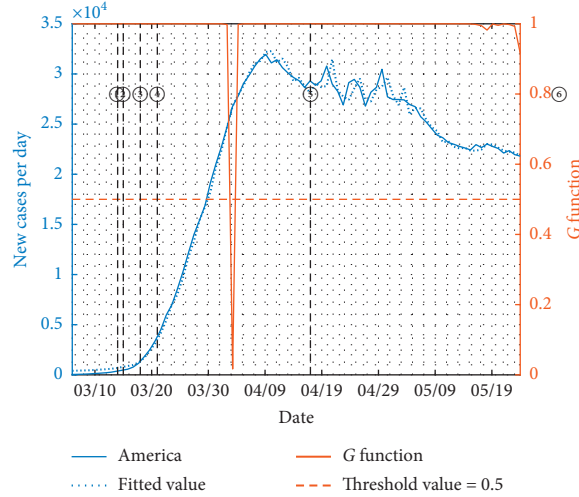


FIGURE 11: The fitting result of America. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① declare state of emergency and close schools, ② impose travel ban to the United Kingdom and Ireland, ③ close the border with Canada, ④ close the border with Mexico, and ⑤ restart the US economy in three stages.

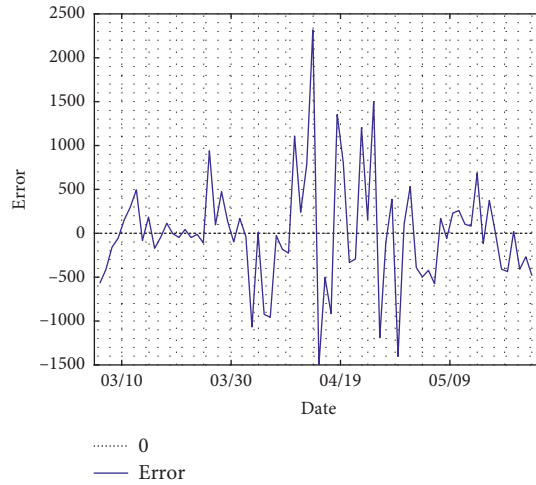


FIGURE 12: The fitting error of America with a mean absolute error of 431.

3.2.3. *Mexico.* The data of daily new confirmed cases from March 21 to July 10, 2020, were selected for Mexico. The moving average is set as 7 to smooth the weekly periodic

fluctuation. The lag orders of autoregression and nonlinear transition are set as $p = 7$ and $d = 4$, respectively. The nonlinear fitting equation is as follows:

$$\begin{aligned} \hat{y}_t = & 217.8 - 0.1y_{t-1} - 0.4y_{t-2} + 1.1y_{t-3} - 1.1y_{t-4} + 1.4y_{t-5} - 1.0y_{t-6} + 1.3y_{t-7} \\ & + (-192.8 + 1.6y_{t-1} + 0.1y_{t-2} - 1.4y_{t-3} + 1.5y_{t-4} - 1.9y_{t-5} + 1.4y_{t-6} - 1.5y_{t-7}) [1 - \exp(-32.0(y_{t-4} - 3308.5)^2)]. \end{aligned} \tag{11}$$

The model fitting results and errors are shown in Figures 17 and 18. The transition speed parameter is $\gamma = 32.0$,

indicating a quick shift in the trend. The boundary for transition is $c = 3308.5$. The mean absolute error is 22.

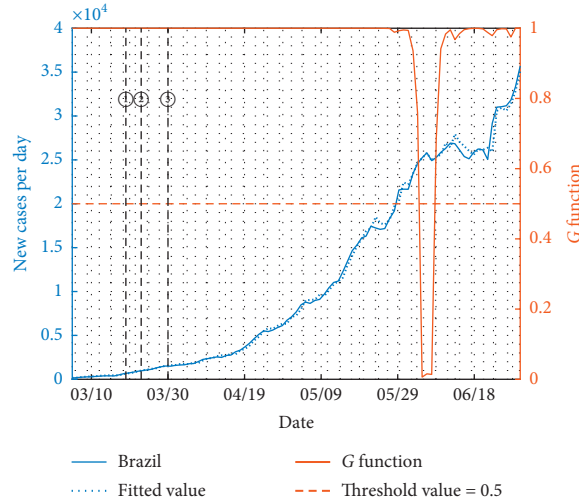


FIGURE 13: The fitting result of Brazil. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① close borders with *Argentina*, *Bolivia*, *Colombia*, *Guyana*, *French Guiana*, *Paraguay*, *Peru*, and *Suriname* within 15 days, ② ban foreigners by air from *China*, *European Union*, *Iceland*, *Norway*, *Switzerland*, the *United Kingdom*, *Australia*, *Japan*, *Malaysia*, and *South Korea*, and ③ implement the lockdown policy.

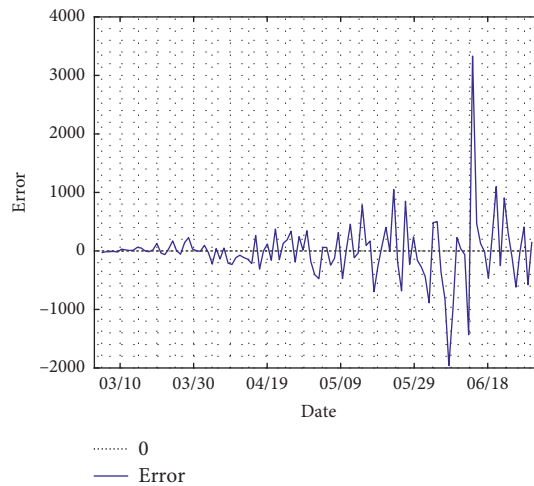


FIGURE 14: The fitting error of Brazil with an average absolute error of 272.

From Figure 17, there is no change for the trend sequence or function G after 3 months since the release of the prevention policy, indicating the ineffectiveness of prevention measures in Mexico. The transition in G in July indicated the slowdown of increase trend in daily new confirmed cases caused by potential natural growth.

3.3. Awareness of the Second Wave. The effective policies in most countries resulted in the temporary control of the COVID-19 pandemic and the reopen of economy and social activities. However, the second wave of the pandemic had

been approaching dramatically. There was no awareness, and it was so fast that no measures had been conducted in time to contain the crazy growth. The pandemic had developed almost naturally without any intervention. In Figures 19 and 20, Italy and Germany are chosen to give an illustration of the gravity of the second wave by the STAR model.

As shown in Figures 19 and 20, the second wave has become severe since September. The G function can only recognize the second wave in Italy due to the steep growth and large amount of daily new confirmed cases. For Germany, the STAR model shows the transitions in both waves

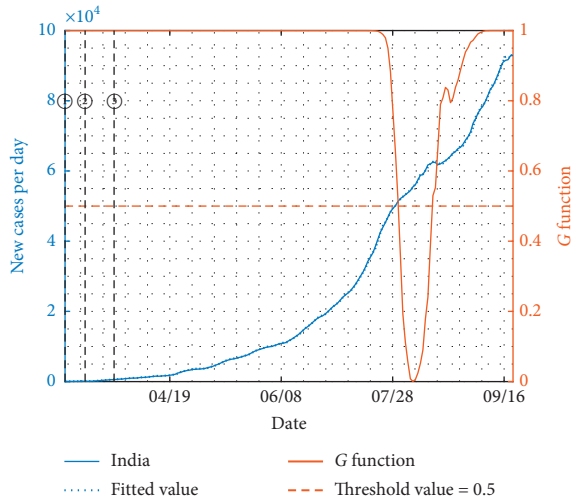


FIGURE 15: The fitting result of India. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① ban travellers from Italy, Iran, Korea, and Japan, ② suspend visas of tourists, and ③ implement a lockdown policy.

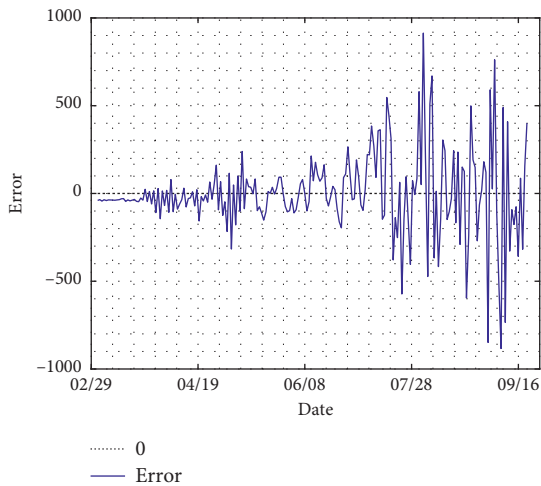


FIGURE 16: The fitting error of India with a mean absolute error of 154.

up to the ending date of our observation, which is October 27. In order to get the second wave under control, stricter measures and more effective policies are urgently needed.

4. Discussion

4.1. Effect of Prevention and Control Policies. We analyzed the daily new cases in 9 typical regions with the STAR model. Combining the release time of policies with the time series of daily new confirmed cases, it can be seen that effective policies would lead to a decrease in the number of daily

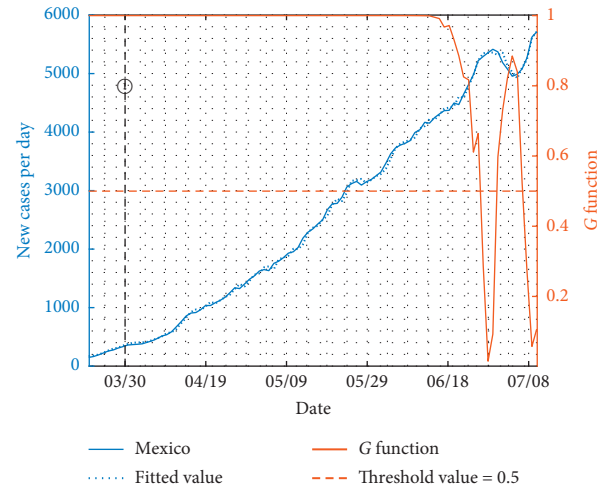


FIGURE 17: The fitting result of Mexico. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of nonlinear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① declare the public health emergency and suspend nonessential activities until April 30.

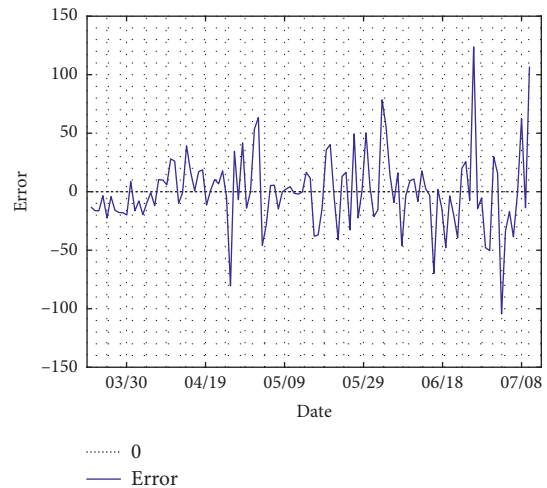


FIGURE 18: The fitting error of Mexico with a mean absolute error of 22.

confirmed cases within 2 weeks to 2 months since the release of certain policies. The inflection points and transitions between different stages of the epidemic curves illustrate the effectiveness of the policy. Meanwhile, in some countries, our model shows the ineffectiveness.

Policies usually include the restrictions on entry of travellers from countries with severe epidemic situations, the closure of national borders, and the suspension of travel. Moreover, schools and nonessential services are closed, large gatherings and public events are banned, and people were asked to stay at home. In the absence of vaccines, these prevention and control policies aiming at blocking transmission are necessary and effective. However, the

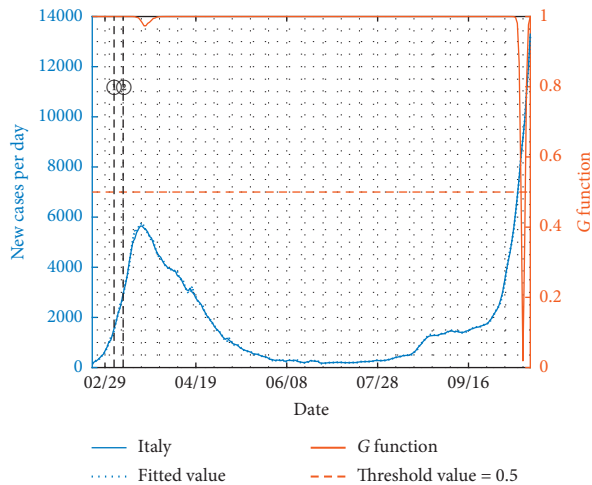


FIGURE 19: The fitting result of Italy containing the second wave. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of non-linear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① close schools across the country and ② nationwide lockdown.

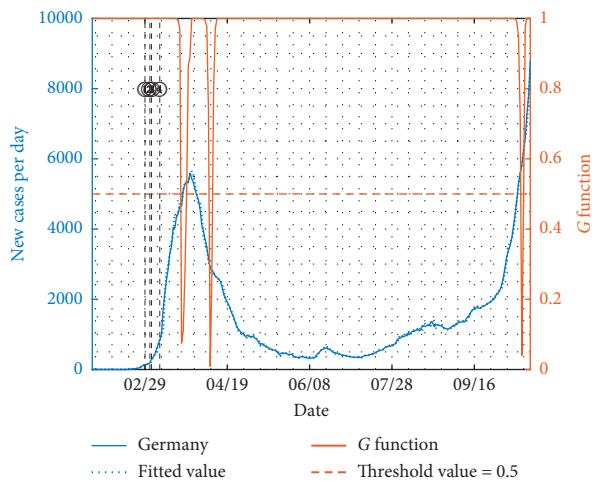


FIGURE 20: The fitting result of Germany containing the second wave. The left axis shows the daily confirmed cases for the real data after smoothness (blue solid line) and its fitted values by the STAR model (blue dashed line). The right axis shows the value of non-linear functions G changing with time (orange solid line) and the reference line with threshold 0.5 (orange dashed line). The vertical dotted line indicates the sequential released policies: ① close schools across the country, ② upgrade the risk level of the outbreak to “high,” and ③ tighten controls on social activities and gatherings.

effectiveness depends on the implementation and cooperation of the masses.

4.2. Be Alert to the Second Wave. We can see from the epidemic curve that many countries are suffering from the second wave of the pandemic, which is more severe than the

first wave. Faced with the social and economic pressure brought by the prevention of COVID-19, it is difficult to achieve a middle ground. Germany, France, UK, and Spain have already upgraded their prevention and containment measures for the current outbreak.

In short, governments should closely monitor the epidemic situation. Moreover, tracing close contacts, large-scale nucleic acid tests, and quarantining of travellers from regions of high risk are ways for preventing the potential pandemic in advance.

Data Availability

Data on daily confirmed COVID-19 cases were collected from the World Health Organization database (<https://www.who.int/data/collections>; <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was jointly supported by the Fundamental Research Funds for the Central Universities (no. 210 2019XD-A11) and the National Natural Science Foundation of China (Grant nos. 11971074, 61671005, 61372109, 61672108, and 61976025).

References

- [1] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, “A novel coronavirus outbreak of global health concern,” *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020.
- [2] J. F.-W. Chan, S. Yuan, K.-H. Kok et al., “A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster,” *The Lancet*, vol. 395, no. 10223, pp. 514–523, 2020.
- [3] A. Lee, “Wuhan novel coronavirus (COVID-19): why global control is challenging?” *Public Health*, vol. 179, pp. A1–A2, 2020.
- [4] Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, and L. A. Meyers, “Serial interval of COVID-19 among publicly reported confirmed cases,” *Emerging Infectious Diseases*, vol. 26, 2020.
- [5] Z. Yu, G. Zhang, Q. Liu, and Z. Lv, “The outbreak assessment and prediction of 2019-nCoV based on time-varying SIR model,” *Journal of University of Electronic Science and Technology of China*, vol. 5, 2020.
- [6] B. F. Maier and D. Brockmann, “Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China,” *Science*, vol. 368, no. 6492, p. 742, 2020.
- [7] L. Hien, K. Veria, K. Piotr et al., “The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China,” *Journal of Travel Medicine*, vol. 27, 2020.
- [8] J. T. Wu, K. Leung, and G. M. Leung, “Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study,” *The Lancet*, vol. 395, 2020.

- [9] J. Hellewell, S. Abbott, G. Amy et al., “Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts,” *Lancet Glob Health*, vol. 8, 2020.
- [10] M. U. G. Kraemer, H. Tian, R. Li et al., “The effect of human mobility and control measures on the COVID-19 epidemic in China,” *Science*, vol. 368, Article ID eabb4218, 2020.
- [11] T. Zhou, Q. Liu, Z. Yang et al., “Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV,” *Journal of Evidence-Based Medicine*, vol. 13, no. 1, pp. 3–7, 2020.
- [12] S. Zhao, Q. Lin, J. Ran et al., “Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak,” *International Journal of Infectious Diseases*, vol. 92, 2020.
- [13] Y. Zhang, “The epidemiological characteristics of an outbreak of 2019 novel coronavirus distases (COVID-19) in China,” *Journal of Epidemiological*, vol. 41, pp. 145–151, 2020.
- [14] W. Guan, Z. Ni, L. Liu et al., “Clinical characteristics of coronavirus disease 2019 in China,” *The New England Journal of Medicine*, vol. 382, 2020.
- [15] G. A. Wellenius, S. Vispute, V. Espinosa et al., “Impacts of statelevel policies on social distancing in the united states using aggregated mobility data during the COVID-19 pandemic,” 2020, <http://arxiv.org/abs/2004.10172>.
- [16] M. Saez, A. Tobias, D. Varga, and M. A. Barceló, “Effectiveness of the measures to flatten the epidemic curve of COVID-19. The case of Spain,” *Science of The Total Environment*, vol. 727, Article ID 138761, 2020.
- [17] N. G. Davies, A. J. Kucharski, R. M. Eggo et al., “Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study,” *Lancet Public Health*, vol. 5, 2020.
- [18] T. P. B. Thu, P. N. H. Ngoc, N. M. Hai, and L. A. Tuan, “Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries,” *Science of The Total Environment*, vol. 742, Article ID 140430, 2020.
- [19] G. Chowell, C. Viboud, L. Simonsen, and S. M. Moghadas, “Characterizing the reproduction number of epidemics with early subexponential growth dynamics,” *Journal of The Royal Society Interface*, vol. 13, no. 123, Article ID 20160659, 2016.
- [20] L. Zhang, “Fitness of the generalized growth to the 2019 novel coronavirus data,” *Journal of University of Electronic Science and Technology of China*, vol. 49, 2020.
- [21] Q. Li, W. Feng, and Y.-H. Quan, “Trend and forecasting of the COVID-19 outbreak in China,” *Journal of Infection*, vol. 80, no. 4, pp. 469–496, 2020.
- [22] R. F. Engle, “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation,” *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.
- [23] J. D. Hamilton, “A new approach to the economic analysis of nonstationary time series and the business cycle,” *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989.
- [24] C. Granger, T. Terasvirta, *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford, UK, 1993.
- [25] K. S. Chan, “On estimating thresholds in autoregressive models,” *Journal of Time*, vol. 7, 1986.

Research Article

The Artificial University: Decision Support for Universities in the COVID-19 Era

Wesley J. Wildman ^{1,2,3} Saikou Y. Diallo ^{1,4} George Hodulik ¹ Andrew Page ^{1,5}
Andreas Tolk ^{1,6} and Neha Gondal ^{1,2,7}

¹Center for Mind and Culture, 566 Commonwealth Ave, Suite M-2, Boston, MA 02215, USA

²Faculty of Computational and Data Sciences, Boston University, 111 Cummington St, Boston, MA 02215, USA

³School of Theology, Boston University, 745 Commonwealth Ave, Boston, MA 02215, USA

⁴Virginia Modeling, Analysis & Simulation Center, Old Dominion University, 1030 University Blvd., Suffolk, VA 23435, USA

⁵Translational Health Research Institute, Western Sydney University, Building 3, David Pilgrim Ave., Campbelltown, NSW 2560, Australia

⁶Modeling, Simulation, Experimentation, and Analytics, The MITRE Corporation, 1001 Research Park Blvd. #220, Charlottesville, VA 22911, USA

⁷Department of Sociology, Boston University, 96 Cummington St, Boston, MA 02215, USA

Correspondence should be addressed to Wesley J. Wildman; wwildman@bu.edu

Received 26 July 2020; Revised 24 November 2020; Accepted 10 December 2020; Published 30 December 2020

Academic Editor: xiaoke xu

Copyright © 2020 Wesley J. Wildman 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.

Operating universities under pandemic conditions is a complex undertaking. The Artificial University (TAU) responds to this need. TAU is a configurable, open-source computer simulation of a university using a contact network based on publicly available information about university classes, residences, and activities. This study evaluates health outcomes for an array of interventions and testing protocols in an artificial university of 6,500 students, faculty, and staff. Findings suggest that physical distancing and centralized contact tracing are most effective at reducing infections, but there is a tipping point for compliance below which physical distancing is less effective. If student compliance is anything short of high, it helps to have separate buildings for quarantining infected students, thereby gracefully increasing compliance. Hybrid in-person and online classes and closing fitness centers do not significantly change cumulative infections but do significantly decrease the number of the infected at any given time, indicating strategies for “flattening the curve” to protect limited resources. Supplementing physical distancing with centralized contact tracing decreases infected individuals by an additional 14%; boosting frequency of testing for student-facing staff yields a further 7% decrease. A trade-off exists between increasing the sheer number of infection tests and targeting testing for key nodes in the contact network (i.e., student-facing staff). There are significant advantages to getting and acting on test results quickly. The costs and benefits to universities of these findings are discussed. Artificial universities can be an important decision support tool for universities, generating useful policy insights into the challenges of operating universities under pandemic conditions.

1. Introduction: The Crying Need for Insight

Operating institutions of higher education under SARS-CoV-2 pandemic conditions is a perilous, complex, and expensive undertaking. Simple simulations of epidemiological models can be adapted to allow university administrators to test combinations of interventions but such models typically neglect the human factors (e.g., social

networks and multiple dimensions of compliance) that heavily influence whether interventions fail or succeed. Bespoke policy simulations incorporating confidential data about students and staff are prohibitive for many schools in terms of their cost and the expertise needed to build them, and they cannot be shared, duplicating effort. Dashboards offering generic advice do not take account of the facts that universities vary widely in what interventions are politically

feasible (e.g., shutting down football is unthinkable in some places, and centralized contact tracing is too controversial in others) or financially achievable (a massive testing, tracing, and quarantine regimen can be prohibitively expensive in terms of materials, space, and personnel). Most dashboards are using available data to support smart extrapolation into the future, while such data are typically not available for universities. Moreover, universities are diverse in terms of what counts as success, and thus they would naturally apply different metrics to evaluate the effectiveness of combinations of interventions (e.g., minimizing infections, or optimizing good outcomes versus bad outcomes). There is an urgent need for flexible and timely policy modeling that can be applied in a host of different institutional contexts.

To respond to this need, the Human Simulation Group created The Artificial University (TAU). TAU is a flexible computer simulation of a university based solely on publicly available web-scraped data about typical class schedules, residence and dining arrangements, age distributions of faculty and staff, on-campus and off-campus activities, gyms, clubs, and commuting. TAU has a threefold configuration process to render the artificial university a close match to any real-world university (Figure 1).

First, TAU can be configured to match university characteristics. TAU handles schools of any size and composition, both two-year and four-year colleges without graduate students and research universities with graduate students. The simulation takes account of faculty and staff age distributions, class schedules, a complex variety of on-campus and off-campus residences with diverse bathroom-sharing arrangements, campus dining halls, specialized quarantine buildings and floors, on-campus gyms and clubs, on-campus sports and public events, and off-campus contacts and commutes. The simulation-based approach permits the application of various knowledge-based prognostication methods and is not limited to smart extrapolation of data.

Second, TAU can be configured to match the range of feasible interventions for a given university setting. Interventions in TAU include allowing or closing gyms, sports, student clubs, and on-campus events; hybrid classes to dedensify classrooms while maintaining live education both in-person and online; testing regimens of varying capacities and reliabilities, both for infection and antibodies, and with varying delays in getting results as well as the possibility of boosting the frequency of infection testing based on antibody testing results, or on whether staff have student-facing jobs; contact tracing of varying intensities, from anonymous apps to centralized tracking by university administrators; and varying intensities of quarantine from self-isolation to placing symptomatic people in quarantine buildings supported by staff. TAU also takes account of compliance with physical distancing requirements, with reporting symptoms to a contact-tracing app, and with self-isolation expectations. These critical human factors can vitiate the effectiveness of interventions.

Third, TAU can be configured to respond to the value-laden perspectives of universities, which yield very different metrics for assessing whether a combination of interventions

is successful. Universities might emphasize the danger of COVID-19 for vulnerable people, including older faculty and staff and people of all ages with preexisting conditions, in which case they will seek to maximize the number of people never infected. Some might prefer a hybrid metric that optimizes the difference between positive outcomes (never infected or recovered) and negative outcomes (deaths or reinfections). Still others might want to take account equity and distributive justice concerns and all will need to account for financial realities. The outcome metrics matter, in the specific sense that they lead to different priorities for intervention arrays. TAU supports all viewpoints on what is an intricately complex, multiattribute, multivalued decision problem. The modeling process itself helps to make problems and solution methods tangible by describing them in the common language of the simulation, so that trade-offs and compromises become clearer and unintended consequences in other value domains may be avoided.

2. Materials and Methods

2.1. Computational Simulation Design. The computational policy simulation that powers TAU is an agent-based epidemiological model (just as for [1]). Each individual agent, whether student, faculty, or staff, moves between states of being susceptible, infected, recovered, and resusceptible, depending in part on biological factors. (It should be noted that these are individual agent states used to trace whether an agent is likely to infect another agent. TAU is not a population-compartmental model, as often used in epidemiology simulation.) Since there have been confirmed reinfections (i.e., the virus causing the second infection is genetically different from the virus causing the first), it is important to allow for the possibility that periods of immunity are short.

TAU is not a spatial model; it is a contact-network model. Network links are the pathways for possible meetings with infectious people. Thus, two agent nodes are linked when there is physical contact and therefore the possibility of infection, indicated by being in class together, living together in a dorm, or both going to the gym or a university event. Links are weighted according to the likelihood of infection. Because universities have a rhythmic schedule over the course of a week, the network is static. Interventions modify network links, reducing the likelihood of meetings, and therefore also of infection. For example, physical distancing, mask-wearing, closing gyms, and dedensifying classrooms by using a hybrid in-person and remote teaching system reduce the probability of infection and thus lighten link weights, or sometimes eliminate links altogether. The higher the levels of compliance, the lower the probabilities of infection. Additional networks can be added, such as family, friends, or other social connections, but, for these networks, usually no data exist on the university level. Further details about TAU's design—including complete documentation of entities, state variables, time scales, networks, network link types, process scheduling, parameters, and initialization—are provided in the online model documentation at <https://github.com/centerformindandculture/TheArtificialUniversity>.

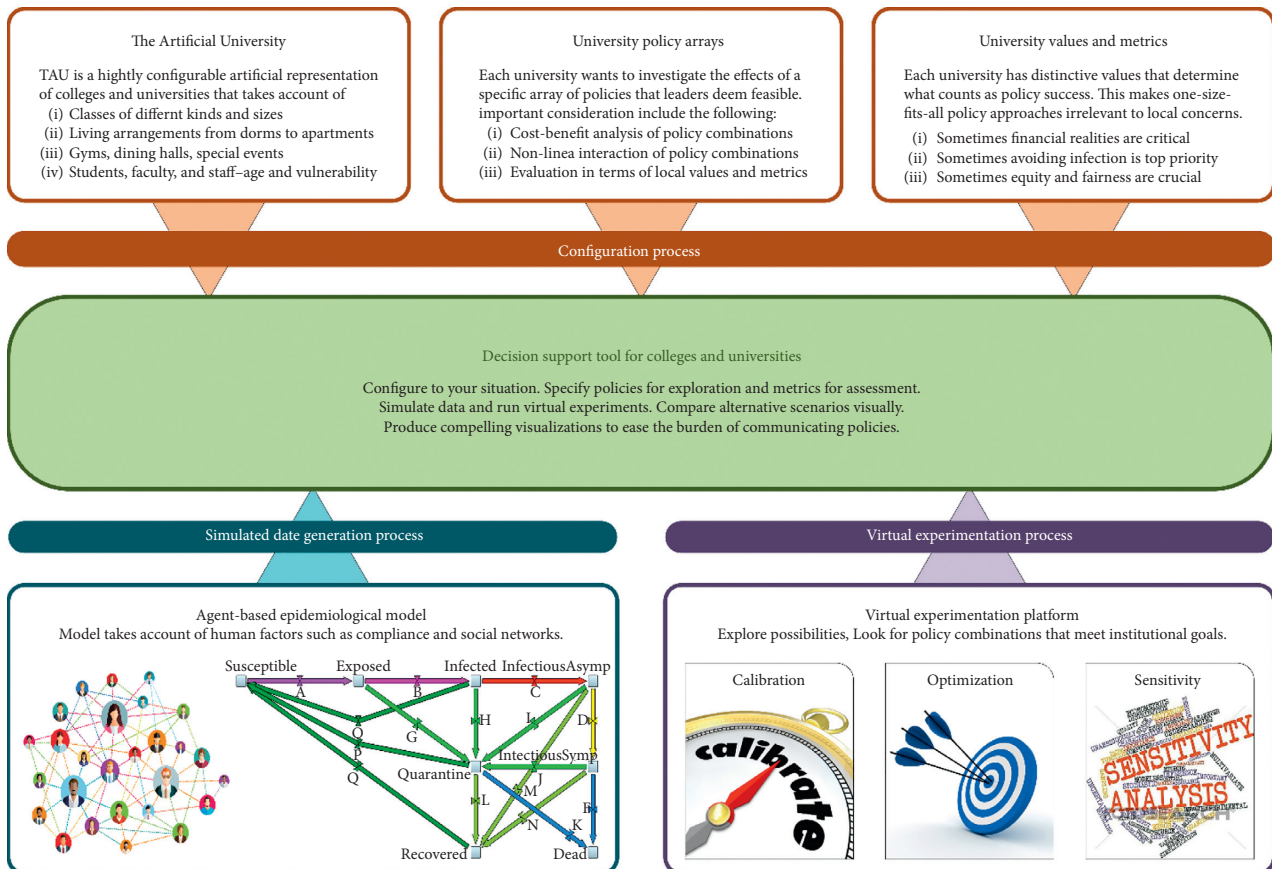


FIGURE 1: Diagram of TAU, a decision support tool for colleges and universities. The threefold configuration process is at the top, the simulation and social network feeding TAU is at the bottom left, and the virtual experimental platform is at the bottom right.

Note that the networks employed in this model are based on copresence (approximating contact on the basis of occupying spaces at the same time) rather than relationship data (social connections such as friendship and familial ties). Copresence data addresses both modeling and practical concerns. With respect to modeling the spread of SARS-CoV-2, sustained copresence (such as seated in close proximity in a classroom) or joint living circumstances (such as suitemates or floor-mates with shared bathroom facilities) are crucial vectors of transmission. Practically speaking, relationship data on ties such as friendship can be highly sensitive to social contexts and, consequently, difficult to generalize. Moreover, relational data are resource intensive whereas copresence data are easier to assume based on the distribution of events and living conditions.

2.2. Design of Experiments: Exploratory Analysis. TAU is highly stochastic and thus we built a large dataset by sweeping the parameter space with 30 replications for each combination of parameter settings, seeking 95% confidence in the outcome variables. This technique of exploratory analysis is often used in areas of deep uncertainty (e.g., [2]). It surfaces emergent behavior and changes in the meta-behavior of the system over the solution space, including tipping points defining the borders of behavior regimes. By

surfacing such information and being clear about assumptions, TAU *promotes thinking deeply into a complex situation* rather than delivering straightforward answers to important questions (see [3]). TAU's capacity to promote deeper understanding of a complex management problem is particularly important given that real-world data for universities does not exist to validate epidemiological models in a quantitative way. Prominent university closures in September 2020 confirm that failures of compliance can comprehensively undermine a COVID-management plan, and this is a useful *qualitative* confirmation of TAU's finding that compliance is the single most important explanatory factor, but quantitative validation must wait for new kinds of datasets to emerge.

We present an evaluation of an array of interventions using health-outcome metrics (e.g., minimizing infections) for two artificial universities, a four-year college and a research university (see Table 1). Most configurable aspects of the universities (e.g., residence arrangements, commuting, class schedules, and activities) are linearly scaled with the population so that the main difference between the two universities is size.

Results for the two artificial universities are similar, so we report here on the smaller university of 6,500 people unless otherwise noted. We do not report on different configurations for universities of the same type and size (e.g.,

TABLE 1: Basic characteristics of the two types of universities tested.

| Characteristic | Four-year college | Large research university |
|----------------------|--|---|
| Size and composition | 6,500 students, faculty, and staff, no graduate students | 37,200 students, graduate students, faculty, and staff |
| Student-facing staff | 10% of staff are student-facing, with 100 interactions per day with students | 10% of staff are student-facing, with 200 interactions per day with students |
| Testing | Maximum 1,000 infection tests and maximum 100 antibody tests per day | Maximum 5,000 infection tests and maximum 500 antibody tests per day |
| Living arrangements | 3 dining halls serving a mix of large dorms sharing bathrooms and apartment style living for on-campus residents; others commute from apartments and homes | 15 dining halls serving a mix of large dorms sharing bathrooms and apartment style living for on-campus residents; others commute from apartments and homes; same ratios as for the four-year college |

Note. Most configurable aspects of the universities are scaled with population size in a nearly linear fashion.

mostly commuter students versus mostly on-campus residents). TAU surfaces surprising and helpful information about interventions and tipping points for their effectiveness, which should be useful to university planners.

We varied parameters related to specific interventions. Using a Latin-hypercube-sampling method, we identified 500 parameter combinations to explore in a parameter sweep. We ran each parameter combination 30 times or until 95% confidence was achieved for three output metrics: the number of people never infected, the total number of infections (including reinfections), and the hybrid metric that measures the difference between positive outcomes (never infected or recovered) and negative outcomes (deaths or reinfections). The resulting 15,000 runs took several days on a machine with a maximum availability of 128 cores and 1 TB RAM. We then ran analyses of the simulated dataset through the statistical package *R*, beginning with a sensitivity analysis of intervention parameters and passing to more specific tests.

We set epidemiological parameters based on reviews of epidemiological and virology literature. For the 6,500-person university reported on here, we fixed a lot of other parameters using findings from scraped data: staff-faculty-student ratios, proportion of students in on-campus dorms, distribution of types of dorms and the number of students sharing bathrooms, distributions of dining halls, frequency of campus events, clubs and sports, class schedules, and off-campus events.

TAU was developed in AnyLogic version 8.5.2. The open-source model and documentation are available at <https://github.com/centerformindandculture/TheArtificialUniversity>. A dashboard for TAU is presented at <http://mindandculture.org/projects/modeling-social-systems/vivid/vivid-dashboard/>. The dashboard facilitates the exploration of the impact of a variety of specific interventions on the university population using two different health-outcome metrics.

3. Results

3.1. Identifying Which Interventions Have the Greatest Impact. The single most important intervention is high compliance with physical distancing. In TAU, physical distancing reduces the probability of an infection through a network link and corresponds in the real world to wearing masks and

keeping physically separate from others. Using the “minimize infections” metric with the adjusted- R^2 test, high compliance with physical distancing explains 70% of the variance (Figure 2).

Adding the variant of contract tracing that involves centralized tracking and strong follow-up to ensure self-isolation brings the total variance explained to 83%. Adding a policy to boost testing frequency of staff with student-facing jobs (e.g., people working in dining halls, cleaning student areas, and meeting intensively with students) further increases the variance explained to 86%.

Note that the adjusted- R^2 test incorporates a penalty for adding additional factors into the regression, so there is a convergence effect as more interventions are included. It follows that the less important interventions in the hierarchy of Figure 2 could still be important when considered alone or in combination with high compliance with physical distancing. To explore this possibility, the TAU dashboard (<http://mindandculture.org/projects/modeling-social-systems/vivid/vivid-dashboard/>) allows users to visualize the projected health effects of varying individual interventions.

If we assume student compliance with social distancing will not be better than 50% (probably a reasonable assumption), do the other interventions still rank order the same way? To answer this question, we fixed faculty and staff compliance of all kinds at a high level and fixed student compliance of all kinds at 50% (save for compliance with forced quarantine in a separate building for infected students, which was fixed to high, and the intervention itself was allowed to be on or off). Under those circumstances, five factors explain the bulk of variance in the “number never infected” metric. The top three hit in the same order as before (after compliance factors are eliminated), though the variance explained is lower, a reminder of how important compliance is as follows:

- (1) Centrally monitor contact tracing (adj- $R^2 = 0.64$)
- (2) Add: boost testing for student-facing staff (adj- $R^2 = 0.75$)
- (3) Add: hybrid classes to dedensify rooms (adj- $R^2 = 0.78$)

The next two factors are as follows:

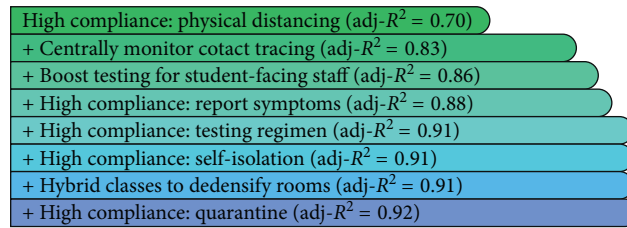


FIGURE 2: Regression-subset diagram for TAU using the “minimize infections” metric on a university of 6,500 people. The single most important intervention is on the top row, explaining 70% of the variance in the outcome metric according to the adjusted- R^2 test. The second row shows the effect of adding the second most important explanatory factor into the regression (adjusted- R^2 lifts to 0.83). As each new factor is added, there are marginal gains in the adjusted- R^2 value.

- (4) Add: forcing quarantine of infected students in a separate building, which improves compliance (adj- $R^2 = 0.78$)
- (5) Add: deprioritizing infection testing for people who test positive for antibodies, freeing tests for others (adj- $R^2 = 0.78$)

Change the metric to “number infected (including re-infections)” and the last two factors switch order.

We also asked what would happen if centralized contact tracing was not used, thereby addressing a big concern for colleges and universities having students, faculty, or staff who are hyperconscious about privacy. In that case, interventions line up in a similar way, with a slightly different order, and much lower adj- R^2 values. Note the appearance of a new factor in fourth place, increasing testing frequency for older faculty and staff:

- (1) Boost infection testing for student-facing staff (adj- $R^2 = 0.31$)
- (2) Add: forcing quarantine of infected students in a separate building, which improves compliance (adj- $R^2 = 0.42$)
- (3) Add: hybrid classes to dedensify rooms (adj- $R^2 = 0.50$)
- (4) Add: boost infection testing for older faculty and staff (adj- $R^2 = 0.52$)

These results drive home the importance of high student compliance with every aspect of the university intervention program, the impact of centralized contact tracing, and the value of “gracefully forcing” infected students into high-compliance mode by using a separate quarantine space.

It is also important to notice which interventions had relatively lower effects on health outcomes. They include the following:

- Closing fitness centers
- Closing student groups and clubs
- Canceling large sport events
- Canceling large events over a threshold size
- Boosting testing frequency for people with health vulnerabilities
- Lengthen contact-tracing history from 7 days to 14 days

In the large research university, TAU produced a similar but not identical list of factors accounting for variance in health outcomes. Here are the top six factors (note the higher values for adj- R^2):

- (1) High compliance: physical distancing (adj- $R^2 = 0.79$)
- (2) Add: boost infection testing for student-facing staff (adj- $R^2 = 0.87$)
- (3) Add: boost infection testing for older faculty and staff (adj- $R^2 = 0.90$)
- (4) Add: centrally monitor contact tracing (adj- $R^2 = 0.93$)
- (5) Add: high compliance: testing regimen (adj- $R^2 = 0.94$)
- (6) Add: high compliance: report symptoms (adj- $R^2 = 0.95$).

3.2. Hybrid Class Structures. With universities facing countless class-action lawsuits alleging failure to provide the promised educational experience due to a switch early in 2020 to remote education, being able to provide a safe, high-quality education that is live and in-person is a priority for university administrators. One way to achieve this is a hybrid system that conducts live classes with two groups simultaneously, one in person and the other remote. This approach dedensifies classrooms while remotely including vulnerable students and students in quarantine. The hybrid system also permits international students who may not be allowed into the country to continue their education.

In TAU, hybrid classrooms work by splitting classes that meet three times a week into three equal-sized, nonoverlapping platoons of students who attend class face to face one day per week and attend remotely for the other two days—all occurring in a room capable of holding the entire class at once. Classes that meet twice per week use two platoons of students, and long classes that meet once per week have two platoons that alternate weeks attending in person. Figure 3 assesses the effectiveness of this intervention strategy. The number never infected by the end of the simulation (Figure 3(a)) is not significantly improved with hybrid classes, but the maximum number infected at any given time (Figure 3(b)) is significantly reduced (by about 30%). It follows that university administrators should not expect

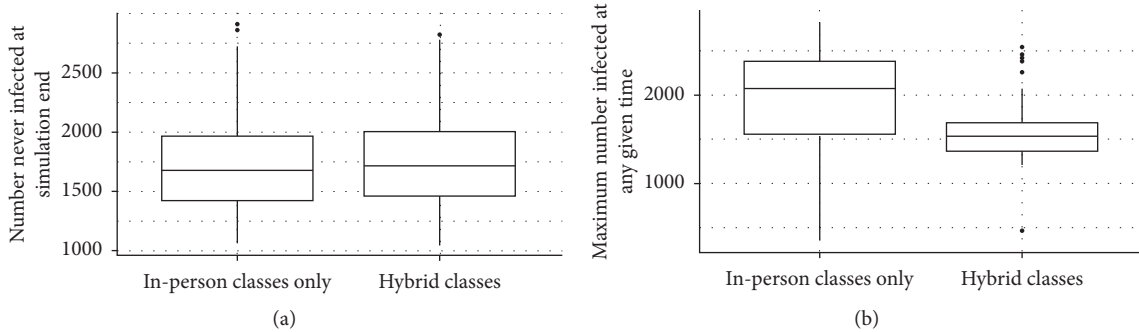


FIGURE 3: Box plots illustrating the effect of introducing hybrid university classrooms. The left image in (a) shows that there is little difference in the number never infected, while the right image in (b) shows that the maximum number infected at any given time is significantly lower with a hybrid class system. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the largest value no further than $1.5 * \text{IQR}$ from the hinge (where IQR is the interquartile range or distance between the first and third quartiles); similarly for the lower whisker. Data beyond the end of the whiskers are outlying points and are plotted individually.

hybrid classes to dramatically lower overall infections, but they can reasonably expect to “flatten the curve” by slowing down the rate of infection.

We hypothesize that platoons in classrooms have a limited impact on increasing the total number of people never infected because most students still live in on-campus housing. It may be possible to select platoons in such a way that they correspond to campus regions, and the resulting spatial compartmentalization might help to confine the spread of any outbreaks. But this is difficult, probably prohibitively difficult, to implement in a real-world class schedule with different class frequencies and compositions, so TAU does not evaluate this possibility.

Hospital emergency departments critically require slowing the rate of infection to avoid overrunning resources. Depending on their policies, universities can face a similar challenge, particularly if they set aside a certain number of buildings with bathroom-equipped rooms for people with symptoms who are quarantined. Hybrid classes can flatten the curve to avoid overrunning those vital and limited resources, even if they only marginally increase the number of people who are never infected. Much the same applies to closing fitness centers, which ought to be vectors of transmission through surfaces and heavy breathing. Such interventions do not appreciably lower the total number eventually infected but they do slow the rate of infection. If flattening the curve is important, then these policies make sense.

3.3. Tipping Points for Compliance with Physical-Distancing Guidelines. Human beings exhibit a rich variety of personalities, convictions, ideologies, and degrees of prosociality, resulting in varying willingness to comply with physical-distancing guidelines. Physical distancing is the single most important intervention for optimizing social-health metrics, but it requires sustained compliance, which young adults in particular often find trying. Consequently, there is an important question about how much of a difference physical distancing really makes.

Figure 4 shows the number of infections at simulation end against the degree of compliance with physical distancing, over all parameter combinations. There is a tipping point around 0.6 (in the simulation, this means 60% probability of being compliant). While the tipping point varies slightly depending on the precise parameter set in play, the response curve always has this shape and indicates that university administrators should expect increased gains with high compliance but few gains with low compliance. That is, if physical distancing is not done well, there is little point in doing it at all. It follows that universities should make physical distancing as easy as possible with masks, spacing marked for queues, and one-way people traffic where possible. University communities (including students themselves) need to promote the benefits of compliance with physical distancing guidelines—particularly with younger people in their community.

3.4. Supplementing Physical Distancing with Testing and Contact Tracing. On the “minimize infections” metric, the next best interventions (after high compliance with physical distancing) are rigorous contact-tracing and flexible testing strategies. How much of a difference do they make on minimizing the number of infections? Findings from TAU suggest the following:

Central tracking (i.e., centralized information about contact tracing, followed by enforcement of isolation among traced individuals) increases the number of people never infected by 14% (this is all of the policy variations with central tracking compared to all of the policy variations without central tracking)

Testing student-facing staff members more often additionally increases the number of people never infected by 7.6%

High compliance with contact-tracing demands, which means reporting when COVID-19 symptoms are experienced or when a test result indicates infection, further increases the number of people never infected by 4.5%

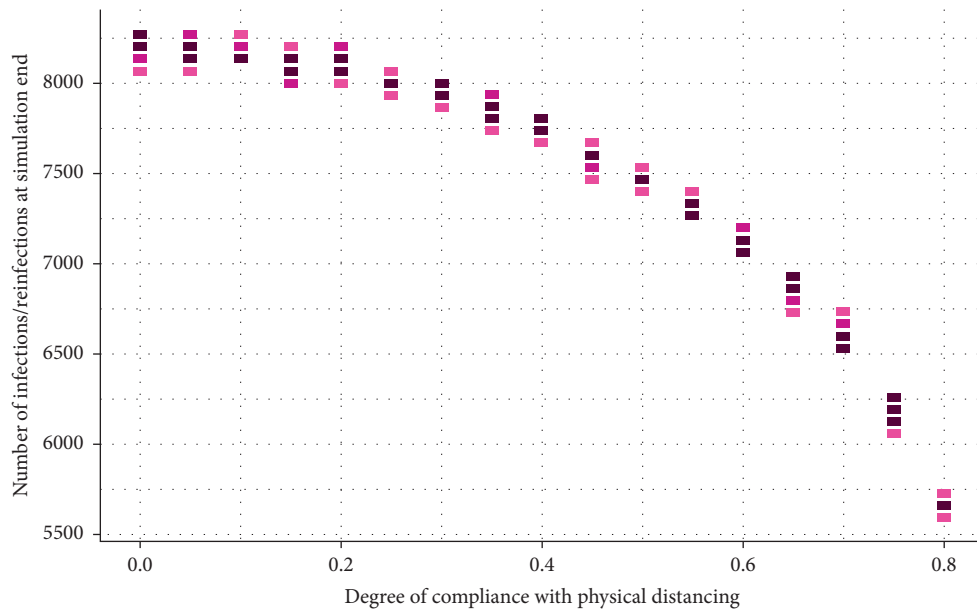


FIGURE 4: The number of infections against the degree of compliance with physical-distancing requirements (including mask-wearing). Note that the “number of infections” metric includes reinfections, so the score can exceed the number of people in the university (6,500). Numerous runs are binned and the color indicates clustering of runs, with darker colors corresponding to a greater number of runs.

Increasing the frequency of testing staff with student-facing jobs poses an important cost-benefit analysis puzzle because testing is expensive and frequent testing is proportionally more costly. We asked TAU how increasing the frequency of testing for staff-with student-facing jobs affects the number of infections (Figure 5). There is an inflection point around a boost of 4 times, taking account of all parameter combinations, suggesting that there is little gain from testing student-facing staff more than about four times more frequently than others.

3.5. Optimizing Testing. TAU examines the scenario where there is no infection testing alongside 16 infection-testing strategies, incorporating an antibody testing strategy. The 16 infection-testing strategies depend on the following:

- Decreasing infection-testing frequency after a positive antibody test (YES if antibody testing is activated, and NO otherwise; positive antibody tests are not possible until at least four weeks after recovery)

- Boosting infection-testing frequency for student-facing staff (YES/NO)

- Boosting infection-testing frequency for people with health vulnerabilities (YES/NO)

- Boosting infection-testing frequency for older people on campus (YES/NO)

There is a delay in receiving infection test results, varying from 24 to 72 hours. Both types of tests vary in cost with the more expensive being more accurate, and there is an economy of scale whereby testing more yields lower per-test cost. TAU is supplied with a fixed number of tests of both types per day, which yields a fixed cost for testing with a specified cost uncertainty (important for universities

standing up internal testing facilities where costs are uncertain). TAU is also supplied with measures of accuracy (the likelihood of false positives and false negatives).

Analysis suggests that any testing regimen is far better than none. For infection testing, the most important factor is the number of tests per week, followed by boosting testing frequency for high-contact nodes in the physical contact network. Other testing options produce marginal returns by comparison. Figure 6 shows the situation in the artificial university after 120 days for six different testing configurations. The horizontal axis shows the average weekly testing frequency for student-facing staff members, the vertical axis shows the number never infected at the simulation end, and each curve shows the number of tests per week. The shape of these curves shows that there is a trade-off between these two considerations such that testing key people more often can be more cost-effective than simply increasing the number of tests.

Figure 7 shows the situation in the artificial university after 120 days for three different testing configurations. The horizontal axis shows the number of viral tests per day, the vertical axis shows the number never infected, the color indicates the testing frequency boost for student-facing staff, and the shape indicates the delay in receiving testing results (24 hours or 72 hours). Getting infection-test results quickly (24 hours rather than 72 hours) makes a significant difference.

3.6. Downstream Consequences of Traced Isolation. In the event of an outbreak on campus, many people will be identified through contact tracing but under most circumstances (assuming proper physical distancing protocols are followed), only a fraction of those individuals will be infected. In some model runs, up to one-third of students needed to self-isolate following contact tracing, despite not being infected.

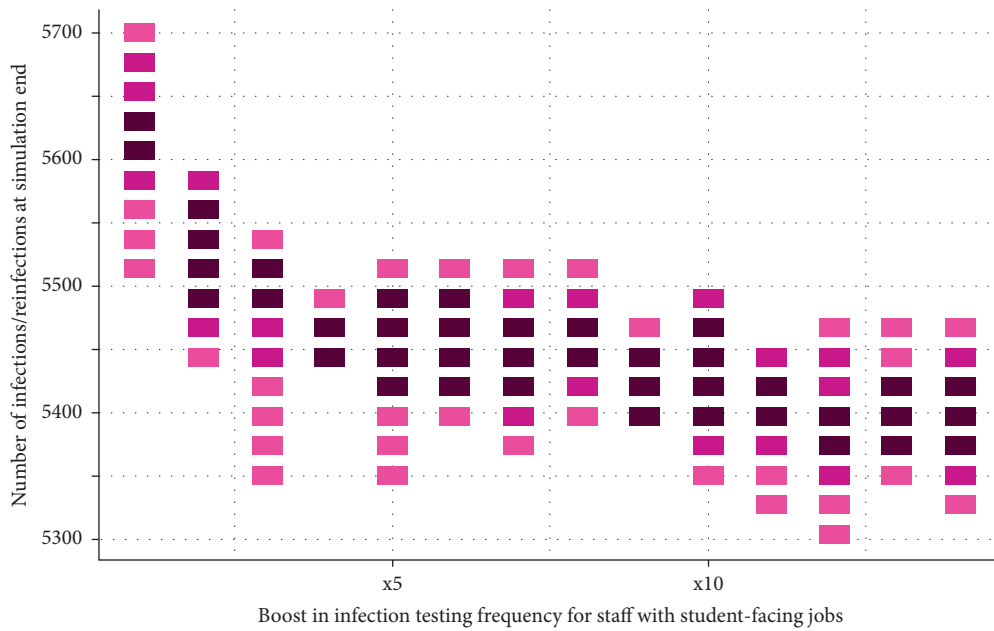


FIGURE 5: The number of infections (or reinfections) against the factor of testing frequency increases for staff with student-facing jobs. Numerous runs are binned and the color indicates clustering of runs, with darker colors corresponding to a greater number of runs.

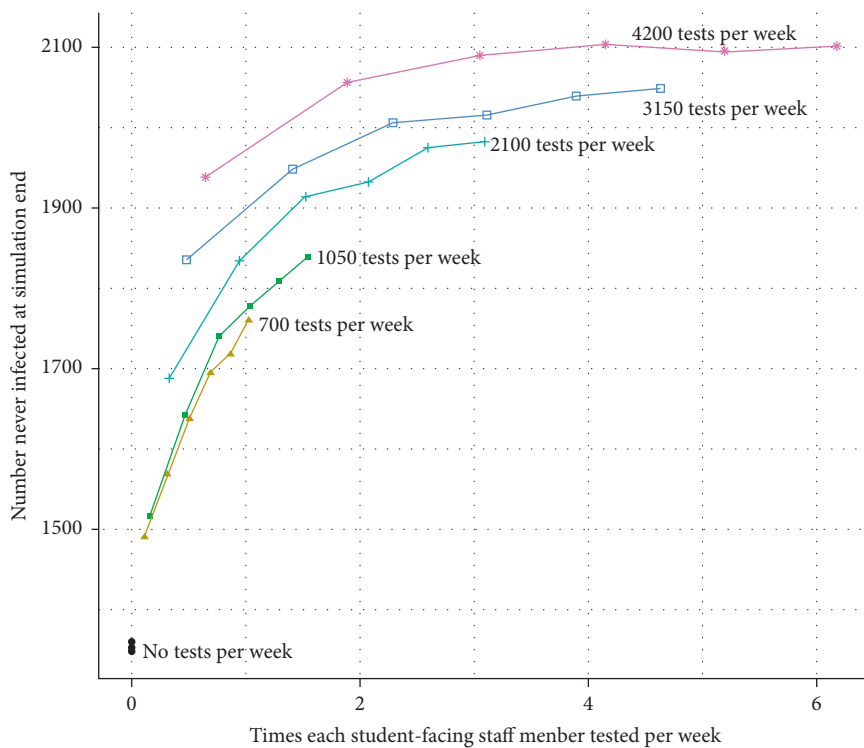


FIGURE 6: Health outcomes for six different testing configurations showing a trade-off between the weekly total number of tests and the testing frequency for student-facing staff. In this figure, each curve is a line of fixed cost, illustrating that better health outcomes can be achieved for lower cost, depending on how testing is performed.

This finding suggests that universities instituting rigorous contact tracing and isolation procedures will have to be ready for large numbers of students in isolation and will need to prepare critical systems accordingly. These include education about the need for self-isolation, food services,

emotional support for people in quarantine (COVID-19 research already shows emotional factors associated with isolation and anxiety are critical for mental health, both acutely and long-term) and infrastructure to facilitate on-going classroom participation remotely.

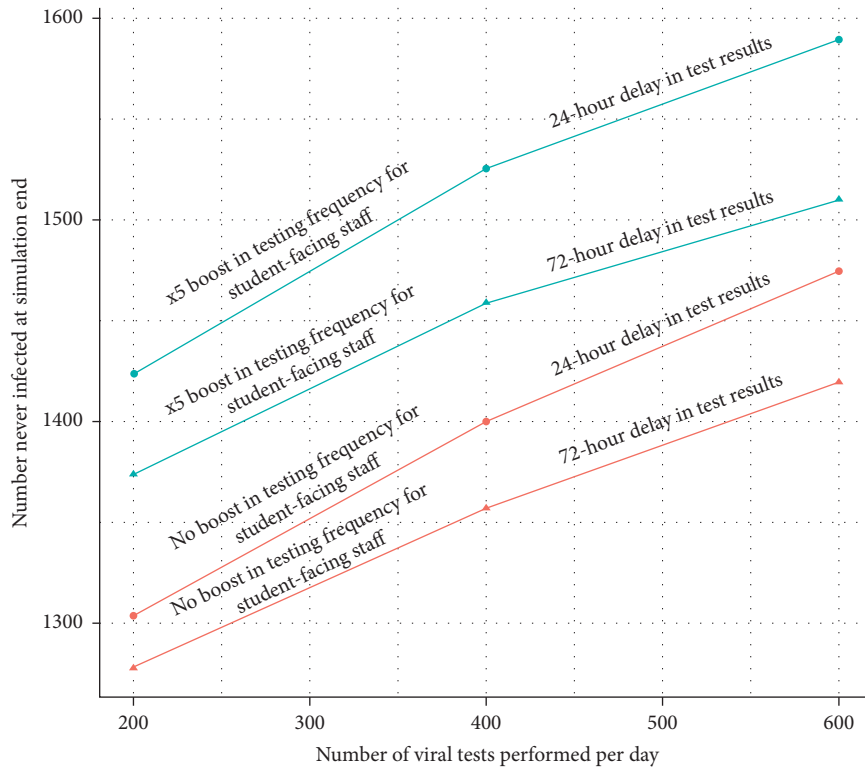


FIGURE 7: Health outcomes for three testing configurations showing the relative importance of increasing testing frequency for student-facing staff and decreasing the delay in getting test results. In this figure, each curve is a line of fixed cost, illustrating that better health outcomes can be achieved for lower cost, depending on how testing is performed and how rapidly results are received and acted on.

Such support systems can be expensive, especially in personnel costs. Therefore, it is also important to keep in mind that traced isolation does not imply infection and that moving trace-flagged students to even more expensive dedicated housing is probably not cost-effective until symptoms appear or testing of self-isolated people shows an infection, at which point people can be moved to avoid spreading the virus through roommates and shared bathrooms.

3.7. Comparing Metrics. Metrics used to assess policy success reflect underlying values, which are critical human factors in a pandemic. To show that metrics matter, we ran a regression-subset analysis for the “minimize infections” metric (Figure 2) and also for the “hybrid” metric that tracks the difference between positive and negative outcomes (described above). Some factors are similarly important for both metrics:

- High compliance with physical distancing
- Hybrid (remote and in-person) classes

Other factors differ in importance:

- Lengthening the history of contact tracing is more important for the hybrid metric
- Boosting frequency of infection testing for student-facing staff is significantly more important on the “minimize infections” metric

It follows that university administrators need to review their values carefully and select the most relevant metrics for their contexts in full awareness that alternative metrics would likely yield different findings for intervention effectiveness.

4. Discussion and Conclusion

Using TAU, we evaluate the possible effects of social distancing, contact tracing, testing, activity closures, dedensifying strategies, and a variety of other interventions. TAU shows that social-distancing requirements (including mask-wearing) have the most significant effect on infections, followed by central tracking and boosting testing frequency for critical networks nodes, which include staff with intensive student contact. However, *high compliance is needed for optimal effect*. For social distancing, we see a tipping point of effectiveness around a compliance rate of 60%, showing the need to create “buy in” among students, staff, and faculty, which calls for targeted publicity campaigns.

The use of a variety of metrics allows us to take different viewpoints into account, promoting a multivalued perspective in which alternatives can be compared and side effects identified. Such multivalued perspectives contest the tendency to focus on one or two domains that capture immediate attention, to the exclusion of others.

We would like to validate TAU against real-world data from universities. Unfortunately, such datasets do not yet exist, though TAU’s finding about the importance of

compliance has been amply confirmed by the fact that the numerous university closures in September 2020 were directly attributable to compliance failures. The New York Times data on universities from September 2020 onwards only report infections and test results and supply no data on policies in place, or on compliance, so cannot be employed to validate TAU.

It is worth thinking further about the role of computational simulations in the absence of the complete real-world datasets needed for comprehensive validation. As noted earlier, even without validation against relevant university data, TAU's architecture, from parameters to processes, are themselves well grounded, so TAU is useful as a way to think deeply into the problem of university management under pandemic conditions, detecting critical explanatory factors and tipping points for intervention effectiveness. This point has been made in a series of important publications, beginning with Troitzsch [4]. More recent practical recommendations for validation can be found in Davis et al. [5], which directly addresses the meaning of simulations like TAU, which are well validated at the low level of causal architecture but cannot be validated at the high level for want of quantitative data of the right kinds. The same point is explored in depth by Saldanha et al. [6].

TAU has limits, which we think of as opportunities for extension. For example, we aim to integrate TAU with county-level data from the COVID-19 Health Care Coalition's dashboard (c19hcc.org) to introduce greater realism in the way TAU handles the porousness of university campuses. We would want to add real-time calibration against university infection statistics. A formal cost-benefit analysis module and an equity-and-justice metric are currently under development for inclusion in the simulation. Additionally, enriching the dashboard at <http://mindandculture.org/projects/modeling-social-systems/vivid/vivid-dashboard/> would simplify the threefold process of configuring TAU, running analyses on TAU-simulated data, and generating visualizations to help communicate policies. There are also a few variations on interventions already included that might prove useful for some universities, such as more intensive contact tracing that attempts to locate superspreaders. In June 2020, we released TAU as an open-source product to allow others to make such adjustments and thereby contribute to the project of helping colleges and universities manage the pandemic.

Despite these limitations, TAU is already a powerful decision support tool for universities. It demonstrates that an artificial university—implemented as an agent-based model using contact networks, integrating an epidemiological model with sensitivity to human factors, and calibrated against publicly available data, following the guidelines of Diallo et al. [7]—can generate valuable insights into the challenge of operating universities under pandemic conditions.

Data Availability

A dashboard interface for exploring TAU is available at <http://mindandculture.org/projects/modeling-social-systems/vivid/vivid-dashboard/>. This site includes links to the model,

documentation, and data (see <https://github.com/centerformindandculture/TheArtificialUniversity>).

Disclosure

The work presented was collaboratively conducted under the COVID-19 Healthcare Coalition. The views, opinions, and/or findings contained in this paper provided by the author are those of The MITRE Corporation and should not be construed as an official government position, or decision, unless designated by other documents. The contributions are approved for public release, distribution unlimited (Case-Nr. 20-01789-7. © 2020 The MITRE Corporation, all rights reserved).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are pleased to acknowledge the work of Dr. Eric Weisel, Executive Director of the Virginia Modeling, Analysis & Simulation Center at Old Dominion University, and some of his staff for their provision of a dashboard interface to support nonprogrammatic access to the simulated data from The Artificial University. This research was funded by the organizations employing the authors: Center for Mind and Culture, Boston University, Virginia Modeling, Analysis & Simulation Center at Old Dominion University, Western Sydney University, and The MITRE Corporation.

References

- [1] P. T. Gressman and J. R. Peck, "Simulating COVID-19 in a university environment," *Mathematical Bioscience*, vol. 328, Article ID 108436, 2020.
- [2] J. H. Kwakkel and P. Erik, "Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty," *Technological Forecasting and Social Change*, vol. 80, no. 3, pp. 419–431, 2013.
- [3] A. F. Siegenfeld, N. N. Taleb, and Y. Bar-Yam, "What models can and cannot tell us about COVID-19," *Proceedings of the National Academy of Sciences*, vol. 117, no. 28, 2020.
- [4] K. G. Troitzsch, "Validating simulation models," in *Proceedings of the 18th European Simulation Multiconference: Networked Simulations and Simulation Networks*, Magdeburg, Germany, June 2004.
- [5] P. K. Davis, A. O'Mahony, T. R. Gulden, O. A. Osoba, and K. Sieck, "Priority challenges for social and behavioral research and its modeling," *Santa Monica, United States: RAND Corporation*, 2018.
- [6] E. Saldanha, L. M. Blaha, A. V. Sathanur, H. Nathan, S. Volkova, and M. Greaves, "Evaluation and validation approaches for simulation of social behavior: challenges and opportunities," in *Social-Behavioral Modeling for Complex Systems*, N. J. Hoboken, Ed., John Wiley & Sons, New Jersey, NY, USA, pp. 495–519, 2019.
- [7] S. Y. Diallo, W. J. Wildman, F. L. Shults, and A. Tolks, "Human simulation: perspectives, insights, and applications," in *New Approaches to the Scientific Study of Religion Series*, Springer, Berlin, Germany, 2019.

Research Article

Risk Assessment of Biological Asset Mortgage Loans of China's New Agricultural Business Entities

Shuzhen Zhu , Yutao Chen , and Wenwen Wang 

Glorious Sun School of Business and Management, Donghua University, Shanghai 20051, China

Correspondence should be addressed to Yutao Chen; 1054580867@qq.com

Received 20 August 2020; Revised 10 October 2020; Accepted 9 November 2020; Published 26 November 2020

Academic Editor: Ye Wu

Copyright © 2020 Shuzhen Zhu 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 large-scale proliferation of China's new type of agricultural entities has given rise to a higher demand for funds. Farmers have insufficient effective collateral, which makes it difficult for them to obtain sufficient loans. Chinese financial institutions have developed a biological asset mortgage loan business to cope with this situation. China has not considered biological mortgages but has been using real estate and asset mortgage models with strong realizability. This innovative financial business has achieved positive results since it was attempted, but it also faces many risks. It is very important to comprehensively and accurately consider the risk factors of biological asset mortgage loans. Based on 1249 production and operation data samples of new agricultural entities in Zhejiang, Henan, and Shandong provinces, this study constructs an XGBoost model for empirical analysis and compares it with logical regression, support vector machine, and random forest algorithms to obtain the optimal model and feature importance value. According to the characteristic importance value, a biological asset mortgage loan risk assessment system with 4 primary indicators and 20 secondary indicators is established, which can effectively identify the biological asset mortgage loan risk of new agricultural entities.

1. Introduction

As China's agricultural production increases in scale, intensification, and specialization, the process of agricultural modernization is gradually accelerating, and the ranks of new agricultural management entities are gradually expanding. Large-scale production requires large amounts of manpower and material resources; consequently, the demand for funds is also gradually increasing. In China, the new type of agricultural management body is defined as an agricultural management organization with a relatively large scale of operation, good material and equipment conditions and management capabilities, high labor production, resource utilization, and land production rate, and includes commercial production as its main objective. According to the People's Bank of China, by the end of March 2019, the balance of various RMB (RMB: Renminbi, currency in circulation in China) loans from financial institutions was 142.11 trillion yuan, up 17.3% year-on-year. The balance of loans from agriculture, forestry, animal husbandry, fishery,

and agriculture in China was 14.68 trillion yuan, up 3.04% from the beginning of the year. Financial institutions' loans to agriculture accounted for 23.72% of total loans, of which loans from pure agriculture, forestry, animal husbandry, and fishery accounted for only 9.68%; this is far from the actual demand for loans from these sectors.

In order to solve the issue of limited funds of new agricultural entities, financial institutions in areas with more developed agricultural industrialization have started to experiment with the biological asset mortgage loan business. Some biological assets have become effective as collateral assets, because of their good liquidity. However, agriculture is a weak industry, and its production and operations are facing great uncertainty, which makes it easier to be affected by natural disasters. Moreover, China's agriculture is undergoing a transition from a small-scale peasant economy to an intensive one. Under the environment of a small-scale peasant economy, the personal credit quality of some farmers is uneven. Coupled with serious information asymmetry and imperfect laws and regulations related to

biological asset mortgage loans, there are numerous potential risks involved in the implementation of these loans. Only some agricultural or financially developed areas carry out this form of business.

This study examines the risk factors involved with biological asset mortgage loans and the main body of new agricultural management, in order to build a risk assessment system to reasonably assess the risk of biological asset mortgage loans. As the magnitude and nature of Chinese agriculture are different from those of other countries, there is less research on establishing a risk assessment system for biological asset mortgage loans in at present, and Chinese scholars often use analytic hierarchy process and fuzzy comprehensive evaluation methods for analysis. These methods are subjective and rely on experts' knowledge and experience, and some risk factors are less available and difficult to apply in practice. Therefore, this study obtains sample data through a questionnaire survey and uses mainstream machine learning methods such as XGBoost, random forest (RF), support vector machine (SVM), and logical regression (LR) to carry out empirical research. According to the value of characteristic importance, this study builds a risk assessment system for biological asset mortgage loans of new agricultural business entities, provides reference for financial institutions to carry out biological asset mortgage loans, and presents several suggestions to promote the development of this innovative financial business and improve the current shortage of funds for new agricultural business entities.

1.1. Literature Review and Related Research. After 5 years of development, China's new agricultural business entities have achieved great progress; however, the overall level of development is low, and there are large differences across regions, which is generally consistent with the development degree of regional economy [1]. Capital, land circulation, market, and talents are the main restricting factors in the development process of new types of agricultural business entities [2–4]. In order to promote the development of new agricultural business entities, financial institutions should innovate financial products and service methods and actively carry out pilot projects such as forest right mortgage, land contractual management right mortgage, and biological asset mortgage loan, in order to activate farmers' assets and form a multichannel financial support system [5, 6].

There are usually basic and agricultural-related parameters in the analysis of agricultural basic credit parameters. The traditional basic credit parameters are still used in general low-level financial institutions and financial underdeveloped countries [7, 8]. Dobbie et al. [9] studied financial and labor market impacts of bad credit reports. They think labor is the critical factor. Sossou [10] found the evidence that farm revenue is positively correlated with land acreage, quantity of labor, and costs of fertilizers and insecticides. In Nigeria, the authors in [11] classified farmers into beneficiaries and nonbeneficiaries. These studies are relatively perfect, but the rapid development of agricultural technology, so that some evaluation criteria have not been effectively analyzed.

Biological asset mortgage loans, a new financial product, has been widely studied by scholars, including their feasibility and necessity, the influencing factors of biological asset mortgage loan willingness, biological asset value evaluation, and biological asset mortgage loan risk research. Biological assets have a clear property right relationship and market value, and do not need to be transferred during the mortgage period, thus resulting in good feasibility [12, 13]. American banks and regulatory agencies think that animal husbandry is the most appropriate collateral [14]. China adopted biological asset mortgage loans relatively late, and only a few regions have attempted it. The research shows that parameters, such as age, total family population, loan experience, production scale, policy cognition, and mortgage financing significance cognition, have significant influence on farmers' willingness to obtain mortgage loans [15, 16]. The value evaluation of biological assets is a necessary process for the implementation of the biological asset mortgage loan business. However, the value of biological assets has always undergone dynamic changes. In addition, the infrastructure of biological asset value evaluation in China is poor, which makes the evaluation more difficult. At present, the international authoritative evaluation methods include the market method, income method, and cost method. Wenjuan et al. [17] compared and analyzed various evaluation methods and considered that the valuation methods of biological assets of different types and growth stages are different. Among crop assets, flower assets should be evaluated through the market method, while other crop assets should be evaluated through the cost method; forest assets with trading market should be evaluated through the market method, as economic forests will bring continuous income. The income method should be used to evaluate the benefits, and the cost method should be used for the evaluation of the young forest assets; the market method should be used to evaluate the livestock assets within an active trading market, and the income method should be used for biological assets that are produced repeatedly. The cost method should be used for the fishery biological assets that are attached to the river water resources and do not exist in the active trading market. However, in different countries and different environments, the scheme of biological asset mortgage needs to be further analyzed and formulated in different regions.

There are many basic approaches to credit risk. Baltensperger [18] first proposed the theory of "credit rationing." The theory holds that compared with applicants with lower credit rating, commercial banks tend to choose applicants with higher credit rating in order to avoid risks and achieve their own business objectives. Altman et al. [19] believe that the logistic model can achieve the best effect in measuring the credit risk of SMEs (small and medium-sized enterprises). Weisen et al. [20] established a credit default risk assessment model for agricultural small and medium-sized enterprises by using the method of logistic combined with factor analysis, and found that the prediction effect of the model is good and has universality. Samad [21] used the probit model to analyze the reasons for the failure of banks, and conducted discriminant analysis on risk factors. It was

proved that the model established by probit can achieve 80% prediction accuracy. Hsieh and Lee [22] used the benchmark model to measure the internal risk of banks. Kamaloo [23] proposed a classifier using immune principles and fuzzy rules to predict quality factors of individuals in banks and other financial institutions. These methods are also commonly used by banks and other financial institutions.

There are few studies on the risk assessment of biological asset mortgage loans, and most of them use an analytic hierarchy process, combined with a fuzzy comprehensive evaluation method [24], to build a risk assessment index system and set weights for the indicators. This type of method is more subjective. Teles [25] compares the SVM and RF algorithms to forecast the recovered value in a credit task. The execution of the projected intelligent systems uses tests and algorithms for authentication of the projected model. Westland [26] developed and tested machine learning models to predict significant credit card fraud in corporate systems using Sarbanes-Oxley (SOX) reports, news reports of breaches, and Fama-French (FF) risk factors. A biological asset mortgage loan risk assessment is similar to a credit risk assessment, which is a binary classification problem. At present, the mainstream method for the study of a dichotomy problem is the machine learning method. Both traditional machine learning methods (such as logical regression and decision tree) or frontier machine learning methods (such as random forest and XGBoost model) can show excellent early warning effects [27–29]. The XGBoost algorithm, with its high classification accuracy, has become the mainstream classification prediction method, and is widely used in various fields such as forecasting and credit rating [30–33]. For unbalanced samples, the XGBoost algorithm can also show better classification performance, which is better than other algorithms.

Considering the imbalance of sample data, this study uses the XGBoost algorithm to fit the possible risk points of biological asset mortgage loans, and compares the classification performance with other algorithms. This study further establishes an evaluation system that can accurately identify the risk of biological asset mortgage loans of new agricultural business entities.

2. Index System Establishment and Research Methods

2.1. Initial Selection of an Indicator System. Considering the risk factors involved before, during, and after the establishment of China's new agricultural business entities, combined with the characteristics of biological assets, and using the 5C factor analysis method of credit analysis, the risks are divided into four categories. The categories are as follows: personal credit risk, specific indicators included in operating risks, biological assets' own risks, and policies and market risks. They are shown in Table 1.

The first category covers the individual credit risk of farmers. The basic characteristics of farmers are used to measure credit risk, and the availability of data is considered. Finally, five secondary indicators are selected to measure gender, age, health status, education level, and asset-liability

ratio. The older the participants, the lower is their risk-taking ability. The more educated the participants, the higher is their willingness to repay their credit. The asset-liability ratio measures the household's repayment ability; the lower the debt to asset ratio, the higher is the repayment ability.

The second category covers the operational risk of farmers. The production and operation conditions determine the output level of biological assets. The output level of biological assets directly affects the income level, which has an impact on the repayment ability and increases the risks of the biological asset mortgage loans. The first-level indicators of business conditions include the following: years of employment, guarantees for others, foreign investment, professional and technical personnel, mechanical automation equipment ownership, simple electronic management, stable and reliable sales channels, stable and high-quality channels of purchase of production materials, grasp of market conditions, agricultural insurance purchases, sales profit margins, and land circulation years.

The third category covers the risk of biological assets, including the difficulty of assessing the value of biological assets, the difficulty of realizing biological assets, the degree of loss caused by natural disasters, and whether the products are high-quality varieties. If the value of biological assets is difficult to determine, the accuracy of the valuation could be reduced; the value of the loans obtained by farmers then exceeds the value of their biological assets, and financial institutions are therefore faced with greater risks. The liquidity of biological assets and the quality of the varieties determine the ability of the product to generate income via sales which will have an impact on the farmers' repayment ability. Natural disasters will affect the growth of biological assets, thus affecting the farmers' income.

The fourth category is policy and market risk, including the range of price fluctuation in the biological asset market, the number of regional asset assessment institutions, and environmental policy risk. The stricter the government's environmental protection requirements, the greater the risk of environmental protection policies affecting farmers. Market conditions are constantly changing due to the influence of factors such as supply and demand, the product cycle, and natural factors; the greater the fluctuation of market prices, the higher is the risk of mortgage loans. The number of institutions in each region determines their asset valuation environment. The better the asset evaluation environment, the more reasonable is the valuation of biological assets, and the lower is the risk associated with biological asset mortgage loans.

2.2. Index System Screening. Through onsite investigations of new agricultural management entities in Zhejiang Province and visiting their production sites, the established index system is screened. Although the term of the land transfer is related to the continuity of operation, the duration of the land transfer contracts signed by the interviewed farmers is generally longer, with them signing these contracts once every few years. Moreover, farmers think that there is almost no sudden interruption of the contract, which will not lead

TABLE 1: Preliminary selection of the risk assessment index system of biological asset mortgage loans.

| Primary indicators | Secondary indicators |
|---|---|
| Basic information relating to the subject | Gender, age, education level, asset-liability ratio, and health status |
| Operating risk | Working years, guarantee for others, foreign investment (part-time), technical personnel, mechanical automation equipment, sales profit margin, stable sales channels, stable source of high-quality means of production, whether there is simple electronic management, the degree of market control, whether to buy agricultural insurance, the number of years of land transfer (In China, land is not owned by individuals. Farmers can get the right to use land from the state. The right of use can be transferred to the farmers through the circulation mode (similar to leasing). However, farmers have the right to recycle), and the expenditure of land transfer |
| Risks of biological assets | Whether the products are high-quality products, the degree of loss caused by natural risks, whether the value of biological assets is easy to evaluate, and the degree of difficulty in realizing biological assets |
| Policy and market risk | The fluctuation range of the product market price, the soundness of asset appraisal institutions, and whether there are environmental risks |

to production interruption. Therefore, this paper will delete the index of “land circulation period”; land transfer expenditure causes production and operation costs. In the survey, it is found that the expenditure of land circulation per mu in almost the same area is the same, basically maintained between 500 and 800 yuan/mu (mu, a Chinese unit of area (=0.0667 hectares), which is not comparable, and the sales profit rate index has already reflected the cost situation. Therefore, the indicator of “land transfer expenditure” is deleted; the index of “the degree of loss caused by natural risk” can reflect the degree of loss caused by of natural disasters, but the research shows that the cost of land circulation is not the same. It is found that vegetables, fruit trees, and other biological assets are planted in greenhouses and are not affected by bad weather. This indicator is difficult to measure and is therefore deleted; in addition, it is difficult to obtain true information relating to health status due to some farmers providing false information; therefore, this indicator is deleted after comprehensive consideration.

The continuous variables were tested using a person test, and the redundant indexes were deleted. There are four continuous variables in the above indicators: age, years of employment in the current industry, profit margin of sales, and asset-liability ratio. After testing the above four continuous variables, it was found that the correlation is not obvious. The correlation coefficient between age and current industry working years is 0.41, the value is less than 0.5, and thus there is no obvious correlation between continuous variables. After screening, a two-tier index system with 21 risk factors is obtained, as shown in Table 2.

Using the experience of previous scholars to measure the risk of biological asset mortgage loans, where farmers have obtained loans from the bank in the past, or there are still outstanding loans, it shows that the bank has conducted an in-depth investigation into these farmers and determined that they have the ability to repay the loans. When these farmers apply for biological asset mortgage loans, the bank determines the farmers’ biology. The risk of a mortgage loan is small; if farmers have not successfully obtained loans or default, it is considered that there is a greater risk of biological asset mortgage loans. Some may not have obtained loans successfully, although the banks have mastered

sufficient information in terms of the main applicant because their production and operation levels and other basic conditions do not meet the requirements of bank lending, as there will be greater risks when applying for biological asset mortgage loans.

2.3. Research Method. There are many related research results on risk assessment methods, such as using the analytic hierarchy process (AHP) and fuzzy comprehensive evaluation methods to set the weight of risk indicators, so as to comprehensively evaluate the risk size, probit regression, and widely used machine learning methods such as support vector machine, random forest, gradient boosting decision tree (GBDT), logistic regression, and XGBoost (extreme gradient boosting). However, existing scholars only use the analytic hierarchy process (AHP) and fuzzy comprehensive evaluation method to evaluate the risk of biological asset mortgage loans. After considering the advantages and disadvantages of various methods, this study uses the current mainstream classification method, namely, XGBoost. The algorithm evaluates the risk of biological asset mortgage loans and compares it with the classification effect of the support vector machine, random forest, and logical regression, and finally obtains an index system that can effectively evaluate the risk of biological asset mortgage loans.

The XGBoost algorithm, proposed by Dr. Chen Tianqi in 2016, is an integrated learning method based on GBDT. It not only retains the original characteristics of GBDT but also greatly improves the training speed and prediction accuracy of the model through a lot of optimization work. It has achieved good results in terms of recommendation, search ranking, user behavior prediction, click-through rate prediction, and product classification. The working principle of the XGBoost algorithm is to establish K cart classification trees. In the running process, the boosting ensemble learning method is used to fit the error of the previous tree with the next tree to reduce the gap between the real value and the predicted value, and shape the model to have the ability to generalize as much as possible. The main feature is to carry out a parallel operation by using a Taylor expansion of the objective function and adding a penalty term, that is, to limit

TABLE 2: Variable definition and descriptive statistics.

| Primary indicators | Secondary indicators | Index description |
|--------------------------------|---|--|
| Subject's personal credit risk | Age | Continuous variable |
| | Gender | Male = 1, female = 0 |
| | Education level | Junior high school and below = 1, high school and technical secondary school = 2, junior college and above = 3 |
| | Assets and liabilities | Continuous variable |
| | Whether to buy agricultural insurance | Yes = 0, no = 1 |
| | Degree of grasp of market conditions | Understand = 1, sure = 2, more sure = 3 |
| | Stable and reliable sales channels | With = 0, without = 1 |
| Main operating risk | Stable and high-quality production material purchase channels | With = 0, without = 1 |
| | Is there simple electronic management? | With = 0, without = 1 |
| | Whether there are professional and technical personnel worker | With = 0, without = 1 |
| | Revenues | Continuous variable |
| | Whether to guarantee others | Yes = 0, no = 1 |
| | Years of agricultural employment | Continuous variable |
| | Invest outward | Yes = 0, no = 1 |
| Risks of biological assets | Is there mechanized automatic equipment? | With = 0, without = 1 |
| | Difficulty of realization of biological assets | Easier = 1, medium = 2, hard = 3 |
| | Is the product of high-quality? | Yes = 0, no = 1 |
| | Is the value of biological assets easy to valuate? | Yes = 0, no = 1 |
| Policy and market risk | Number of regional asset appraisal institutions | Less = 1, more = 0 |
| | Biological asset market price fluctuation range | Small = 1, medium = 2, big = 3 |
| | Is there any environmental policy risk? | With = 0, without = 1 |
| Result | Is there a greater risk? | Yes = 0, no = 1 |

the number of leaf nodes to prevent the model from being too complex. The objective function of the model is composed of the loss function and regularization term:

$$\text{Obj}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \Omega(f_t), \quad (1)$$

where $l(y_i, \hat{y}_i^{(t)})$ is the difference between the real value and the predicted value and $\Omega(f_t)$ is the regularization term, which limits the number of leaf nodes.

2.4. Data Sources. The samples in this study come from new agricultural business entities in 14 regions in the three provinces of Zhejiang, Shandong, and Henan, including Hangzhou, Ningbo, Wenzhou, Shaoxing, Jiaying, Taizhou, Jinhua, Huzhou, Quzhou, Lishui, Zhoushan, Zaozhuang, Weifang, and Xinyang. According to the established biological asset mortgage loan risk indicators, the questionnaire was designed and conducted during on-site investigations. Accompanied by relevant personnel of the regional banks, the new agricultural business entities were interviewed, and the questionnaires were issued to obtain first-hand information. The investigation obtained 1498 data points.

3. Analysis and Results

3.1. Data Preprocessing and Data Descriptive Statistical Analysis. The data in this article come from field surveys and contain some problems such as missing and distorted data.

Subtype variables should be numerically processed when using machine learning modeling, and continuous variables should be normalized when using logistic regression and support vector machine modeling to avoid inaccurate results caused by large data values. Therefore, this study performs one-hot encoding on the categorical data and performs missing values, outliers, and normalization at the same time.

3.1.1. Outlier Handling. In terms of continuous variables, it is impossible to judge whether the data are caused by human error or whether the data themselves are true; therefore, outliers cannot be deleted directly. An excel filtering operation is used to view the distribution of data and subjectively determine whether to delete or keep unreasonable data. For example, for the sales profit rate feature, the maximum value of the sales profit rate itself is 1; if the data value exceeds 1, there is the possibility of incorrect filling. Deleting such sample data has little effect on the training of the model and can therefore be deleted directly.

3.1.2. Missing Value Processing. In order to ensure the integrity of the data, the missing data need to be processed, and the processing of missing values is determined according to the degree of missing data. This study uses 20% as the limit, deleting data with missing values over 20%, and fills in data with missing values less than or equal to 20%. Among them, continuous data are filled with averages, and subtype data are filled with modes.

3.1.3. One-Hot Encoding. The working principle of one-hot encoding is to use N-bit status registers to encode N states. Each state has its own independent register bit, and at any time, only one of them is valid. One-hot encoding can, to a certain extent, expand the features to prevent overfitting of the model. If the sample contains categorical data, such as gender, it usually uses 0,1 as a label to mark the value of the variable, where there is no logical relationship in itself. However, when training the model, the computer will default to a numerical value and make logical judgments based on the size of the data. Therefore, in order to facilitate the input of the model, discrete variables need to be quantified into numerical values. This article performs one-hot coding on 17 variables including gender, education level, whether to invest abroad, whether to guarantee others, and whether there are professional and technical personnel. Finally, model training and prediction data can be carried out directly.

3.1.4. Normalization. Normalization processing, refers to the processing of numerical data to limit it to a certain range, so as to prevent certain characteristics from being “eaten” by big data due to their own expression problems. This results in each feature being treated fairly by the classifier. When using algorithms to test data, the tree-based algorithm model does not need to normalize the data, while the support vector machine and logistic regression need to normalize the data before the experiment is conducted, placing the data on a different dimension. In this study, the maximum-minimum method is used to process the data so that the numerical data are in the interval [0, 1]. The specific formula is as follows:

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (2)$$

where X is the specific value of the selected variable, x_{\max} is the maximum value that the variable can obtain, and x_{\min} is the minimum value that the variable can obtain.

3.1.5. Data Descriptive Statistical Analysis. After processing the data, a total of 1249 samples for modeling were finally obtained, with 21 feature numbers, including 1,038 samples from Zhejiang Province, 116 from Henan Province, and 95 from Shandong Province. The biological assets involved include field crops, live animals, and economic crops, a total of 39 products. The specific products are listed in Table 3.

Python is used to perform statistics on each feature to obtain the distribution of each feature value, such as the average value, standard deviation, maximum value, and minimum value. See Table 4 for details. From the statistical results, the majority of the sample is male, accounting for about 86% of the total sample; the average age of the new agricultural business entities is 46.8 years old, the oldest in the sample is 67 years old, while the youngest is only 23 years old. Some young people have joined large-scale agricultural productions, but their age is generally high at present. The average education level is 1.6, indicating that the education level of farmers is generally not high, but there are also some high-intellectuals engaged in agricultural production. The

average working life of farmers is 12–13 years. Long-term agricultural production has allowed it to accumulate a lot of experience; the average asset-liability ratio is 36.23%; 28.87% of farmers provide guarantees for others; 21.22% of farmers also invest in other industries while engaged in agriculture; nearly half of them are equipped with professional and technical personnel. As they have been engaged in agriculture for many years, some business entities are more skilled in agricultural production and can be used as professional and technical personnel. Most farmers have introduced mechanized production, and a small number of farmers have simple electronic management. Most entities have a good grasp of market conditions and have more accurate judgments in terms of their environment, resulting in an average sales profit rate of 29%. In order to avoid natural risks, approximately half of the entities have purchased agricultural insurance, and most farmers produce ordinary products. Approximately, 13% of the main products produce good varieties, and the price fluctuations of the products produced are small. The balance of supply and demand can be achieved in the domestic market. Evaluation of the value of some biological assets is difficult. At the same time, the environmental protection policy risk faced by the operating entity, and the risk of production interruption are relatively small.

3.2. Construction of Risk Assessment Model of Biological Asset Mortgage Loan Based on XGBoost Algorithm. The XGBoost algorithm is used as a modeling tool and compared with the classification performance of LR, SVM, and RF algorithms. After optimizing the parameters of the model, the best model and importance value of each feature of the biological asset mortgage loan risk assessment are obtained.

3.2.1. Model Performance Evaluation Index. The confusion matrix (Table 5) is the most basic, intuitive, and easiest way to measure the accuracy of model classification. From the confusion matrix results, the accuracy, recall, F1-Score, classification accuracy of positive samples, and classification accuracy of negative samples are obtained. Rate and other evaluation indicators are as follows.

In addition to the results of the confusion matrix, a receiver operating characteristic (ROC) curve and area under the curve (AUC) values are also important evaluation indicators. The ROC curve describes the relationship between the model's TPR and FPR, where TPR is the proportion of the number of positive samples that are correctly classified by the model to the total number of positive samples, and FPR is the number of negative samples that are incorrectly classified by the model to the total number of negative samples. The horizontal axis of the ROC curve represents the FPR, and the vertical axis represents the TPR. For classification problems, one predicts a positive sample probability for each sample and then compares it with the set threshold to determine whether it is a positive sample or a negative sample. The AUC value is the area under the ROC curve, and the value range is [0,1]. The larger the AUC value, the better the model classification effect. The purpose of the

TABLE 3: Classification of sample products.

| Livestock products | Fruits | Vegetables | Food crops | Seedlings and flowers |
|--------------------|------------|---------------|------------|-------------------------------|
| Pig | Pomelo | Toona | | Ginkgo |
| Chicken | Tangerine | Soy | | Flowers |
| Duck | Grape | Eggplant | | Bonsai |
| Beef cattle | Pitaya | Tomato | | Succulent plants |
| Mutton | Loquat | Mushroom | | Chrysanthemum |
| Goose | Peach | Zizania | Rice | Camphor tree |
| | Pear | Black fungus | | Chinese herbal medicine crops |
| | Persimmon | Lily | | Red plum |
| | Kiwi | | | Pecans |
| Mink | Watermelon | Bamboo shoots | | Green maple |
| | Fig | | | |
| | Strawberry | | | |

TABLE 4: Data descriptive statistical analysis.

| Index | Mean | Standard deviation | Minimum | Maximum |
|---|-------|--------------------|---------|---------|
| Age | 46.82 | 8.08 | 23 | 67 |
| Gender | 0.14 | 0.35 | 0 | 1 |
| Education level | 1.61 | 0.75 | 1 | 3 |
| Assets and liabilities | 0.36 | 0.65 | 0 | 12.5 |
| Whether to buy agricultural insurance | 0.54 | 0.50 | 0 | 1 |
| Degree of grasp of market conditions | 2.34 | 0.58 | 1 | 3 |
| Stable and reliable sales channels | 0.20 | 0.40 | 0 | 1 |
| Stable and high-quality production material purchase channels | 0.32 | 0.47 | 0 | 1 |
| Whether there is simple electronic management | 0.67 | 0.47 | 0 | 1 |
| Whether there are professional and technical personnel worker | 0.49 | 0.50 | 0 | 1 |
| Revenues | 0.29 | 0.21 | -1 | 0.9 |
| Whether to provide guarantee for others | 0.71 | 0.45 | 0 | 1 |
| Years of agricultural employment | 12.13 | 8.11 | 1 | 43 |
| Invested outward | 0.79 | 0.41 | 0 | 1 |
| Is there mechanized automatic equipment? | 0.61 | 0.49 | 0 | 1 |
| Difficulty of realization of biological assets | 2.21 | 0.77 | 1 | 3 |
| Is the product of high-quality? | 0.87 | 0.33 | 0 | 1 |
| Is the value of the biological assets easy to evaluate? | 0.32 | 0.47 | 0 | 1 |
| Number of regional asset appraisal institutions | 0.37 | 0.48 | 0 | 1 |
| Biological asset market price fluctuation range | 1.40 | 0.75 | 1 | 3 |
| Whether there is any environmental policy risk | 0.70 | 0.46 | 0 | 1 |
| Whether there is a risk | 0.25 | 0.43 | 0 | 1 |

risk assessment model is to be able to identify users with higher risks; thus, the correct rate of the negative sample classification is very important. The positive and negative samples in this paper are not balanced; therefore, the AUC value and the classification accuracy of negative samples are mainly used as the model performance evaluation indicators.

3.2.2. Model Construction and Parameter Optimization.

Using the Python 3.8 series as the modeling tool, install the XGBoost library and import the data, divide the data by the setout method, and randomly select 75% of the data as the training set through the frac in the df.sample function, and the remaining 25% of the test set. The distribution of the positive and negative samples of the training set and the test set is relatively balanced, as shown in Table 6.

The selection of XGBoost model parameters has a great impact on the performance of the model. The parameters included in XGBoost are divided into three types: general parameters, tree booster parameters, and learning task parameters. The learning rate parameter eta and row-sampling parameters in the tree booster parameters, such as sub-sample, column-sampling parameter colsample_bytree, and L1 and L2 regularization item weights alpha and lambda play a key role in preventing model overfitting. For the binary classification problem, the objective function parameter will be set to binary: logistic and the scale_pos_weight parameter will be set to 0.34 according to the positive and negative ratios of the sample.

We train the model; except for the fact that the maximum tree depth is selected as 3, the other parameters are all the default values of the XGBoost algorithm. The trained model is tested on the test set, and the test set evaluation

TABLE 5: Confusion matrix.

| | | Predictive value | |
|--------------|---|---------------------|---------------------|
| | | 0 | 1 |
| Actual value | 0 | True negative (TN) | False positive (FP) |
| | 1 | False negative (FN) | True positive (TP) |

TABLE 6: Training set and test set sample distribution.

| | 0 (risky sample) | 1 (low-risk sample) |
|--------------|------------------|---------------------|
| Training set | 237 | 700 |
| Test set | 79 | 233 |

indicators are obtained. The model accuracy rate is 93.81%, the recall rate is 90.99%, the F1-score is 92.37%, the accuracy rate of negative samples reaches 82.28%, and the accuracy rate is 90.99% in terms of classification of positive samples. The accuracy rate is 90.99%, the AUC value is 0.9363, and the classification performance of the model is better.

Thereafter, the parameters are optimized. Hyperparameter optimization methods include grid search, random search, and Bayesian optimization. Grid search first specifies a subset of the hyperparameter space, exhausts all combinations of the given hyperparameters, and tries to find a set of optimal hyperparameters. In scikit-learn, the grid search can start from the parameter `param_grid`. The specified parameter grid is exhaustively obtained to obtain the optimal parameters to realize the grid search, A random search of the parameter space is conducted, and the value of the parameter is obtained through a probability distribution extraction. A grid search is suitable for small datasets and is currently the most widely used method in hyperparameter optimization. Therefore, this study selects a grid search as the parameter optimization method, a nested cross-validation in the grid search, and uses a three-fold cross-validation to find the optimal number of decision trees, and then grid search for other parameters step by step. There are eight parameters selected for optimization: `eta`, `max_depth`, `min_child_weight`, `gamma`, `subsample`, `colsample_bytree`, `reg_alpha`, and `reg_lambda`.

When tuning parameters, if the learning rate `eta` is too large, it will easily fall into overfitting. Being a too small value will increase the number of trees and increase the amount calculated for model training. Therefore, we first select a higher `eta` and set it to 0.1. Then, the cross-validation function in the XGBoost function is used to select the number of decision trees with the optimal `eta` value, tuning each parameter in turn, then reselect the `eta` value, and use the AUC value as the model performance evaluation index to finally obtain the best XGBoost model and optimal parameter combination. The optimal parameter combination is shown in Table 7.

Under the optimal parameter combination, the AUC value on the test set is 0.9493, and the ROC curve (Figure 1) is steep. The results of the confusion matrix are shown in Table 8. According to the confusion matrix, the accuracy of the negative sample classification of the model is 83.54%, the accuracy of the model is 94.17%. The recall rate is 90.13%,

and the f1 score is 92.11%, which is generally better than that of the model before tuning. The classification performance is better.

3.2.3. A Comparison of the Classification Performance of Different Machine Learning Methods. Select LR, SVM, and RF algorithms for comparison, use the default parameters of each algorithm to train on the same dataset, and compare the evaluation index values after testing on the same test set. Then, optimize the parameters of each algorithm. The grid search method based on a three-fold cross-validation is still used in parameter optimization, and the model is retrained after finding the optimal parameter combination. The prediction is made on the test set, and the classification performance indicators such as the classification accuracy of negative samples and the AUC value are compared. Finally, the classification index values before and after optimization of each model are shown in Table 9.

From the comparison of the above four models, it can be seen that the classification performance of the XGBoost is better than that of other models before parameter tuning, which also confirms the desirability of choosing the XGBoost model to evaluate the risk of biological asset mortgage loans in this study. After parameter optimization, the negative sample classification accuracy rate of the XGBoost model is as high as 83.54%, which is much higher than that of the other models. Therefore, the XGBoost model is more suitable for biological asset mortgage risk assessment because of its higher classification accuracy and AUC value.

3.2.4. Analysis of Empirical Results. Based on the above empirical analysis, this study establishes a biological asset mortgage loan risk assessment model based on the XGBoost, which has a good classification effect. According to the established XGBoost model, the importance value of each feature is obtained with the `xgb.feature_importances_` command. There are three calculation methods for feature importance: "weight," "gain," and "cover." In this paper, the default method "weight," is the number of times that features are divided into attributes in all the trees. This is calculated as the standard, and the importance of each feature is shown in Figure 2.

It can be seen from the feature importance value that, except for the feature "whether the product is a high-quality product," where the importance value is 0, other features have a certain impact on the risk of biological asset mortgage loans. This shows that there is no significant relationship between product variety and loan risk. The difference between high-quality products and ordinary varieties is that the final circulation market is different, and the products can sell well in their respective circulation markets. Summarizing the above characteristic importance values, the established biological asset mortgage risk assessment system for new agricultural business entities is shown in Table 10.

TABLE 7: Optimal parameter combination.

| Hyperparameter | The optimal value |
|------------------|-------------------|
| Learning_rate | 0.1 |
| Max_depth | 3 |
| Gamma | 0.06 |
| Min_child_weight | 0.5 |
| Subsample | 0.8 |
| Colsample_bytree | 0.8 |
| Reg_alpha | 0.9 |
| Reg_lambda | 1.5 |

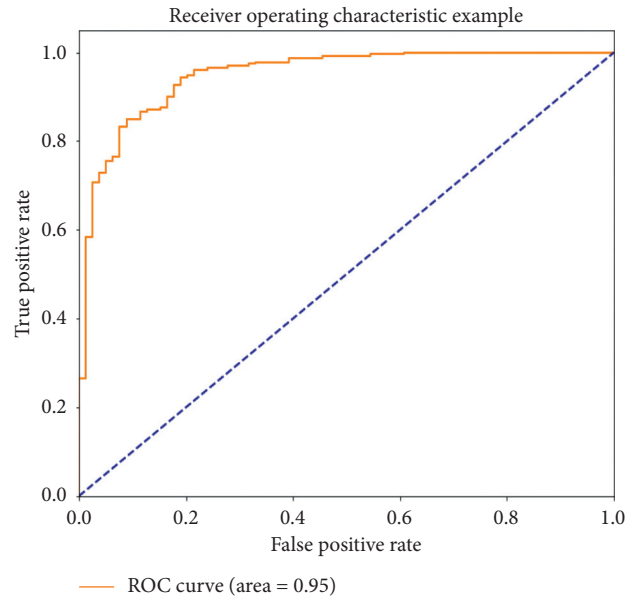


FIGURE 1: ROC curve of the XGBoost model after optimization.

TABLE 8: Confusion matrix results of the XGBoost model after tuning.

| | | Predictive value | |
|--------------|----------------------|-------------------|----------------------|
| | | 0 (risky farmers) | 1 (low-risk farmers) |
| Actual value | 0 (risky farmers) | 66 | 23 |
| | 1 (low-risk farmers) | 13 | 210 |

TABLE 9: Classification index value of each model before and after tuning.

| Model | AUC (%) | Negative class accuracy (%) | Positive accuracy (%) | F1-score (%) |
|-------------------------------|---------|-----------------------------|-----------------------|--------------|
| Before parameter optimization | | | | |
| XGBoost | 93.63 | 82.28 | 90.99 | 92.37 |
| LR | 82.84 | 54.43 | 92.27 | 88.84 |
| SVM | | 16.46 | 98.71 | 86.96 |
| RF | 92.88 | 60.76 | 96.14 | 91.80 |
| After parameter optimization | | | | |
| XGBoost | 94.93 | 83.54 | 90.13 | 92.11 |
| LR | 82.82 | 50.63 | 93.13 | 88.75 |
| SVM | | 67.09 | 93.13 | 91.18 |
| RF | 95.31 | 68.35 | 97.42 | 93.61 |

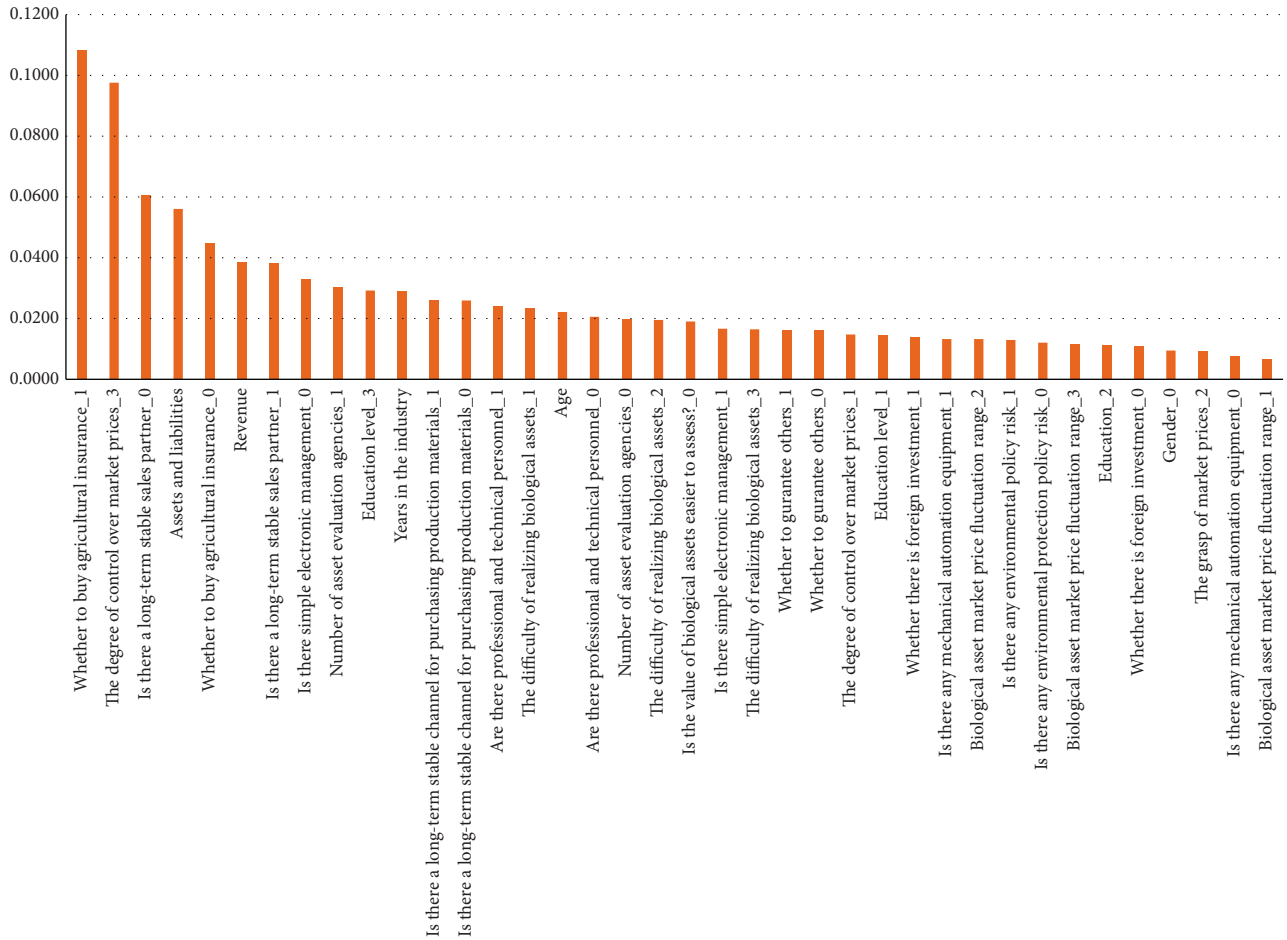


FIGURE 2: The importance of characteristics of biological asset mortgage risk assessment.

TABLE 10: The risk assessment index system of biological asset mortgage loans for new agricultural business entities.

| Primary indicators | Secondary indicators | Index importance value |
|--------------------------------|---|------------------------|
| Subject's personal credit risk | Assets and liabilities | 0.0561 |
| | Education level | 0.0559 |
| | Age | 0.0222 |
| | Gender | 0.0097 |
| | Whether to buy agricultural insurance | 0.1537 |
| Main operating risk | Degree of grasp of market conditions | 0.1224 |
| | Stable and reliable sales channels | 0.0990 |
| | Stable and high-quality production material purchase channels | 0.0525 |
| | Whether there is simple electronic management | 0.0501 |
| | Whether there are professional and technical personnel worker | 0.0450 |
| | Revenues | 0.0386 |
| | Whether to guarantee others | 0.0329 |
| | Years of agricultural employment | 0.0291 |
| | Invest outward | 0.0255 |
| | Whether there is mechanized automatic equipment | 0.0212 |
| Risks of biological assets | Difficulty of realization of biological assets | 0.0600 |
| | Is the value of biological assets easy to evaluate? | 0.0192 |
| Policy risk and market risk | Number of regional asset appraisal institutions | 0.0503 |
| | Biological asset market price fluctuation range | 0.0316 |
| | Whether there is any environmental policy risk | 0.0246 |

4. Conclusions

4.1. Basic Conclusion. The operating status of the lender is the most important type of risk. Among them, “whether to buy insurance” is an important indicator to measure the risk of biological asset mortgage loans for new agricultural operators, and this indicator can be used as the “threshold” for financial institutions to issue biological asset mortgage loans. Furthermore, the new agricultural business entities’ grasp of market conditions and whether the products have reliable sales channels are also key indicators that will trigger the risk of biological asset mortgage loans. The asset-liability ratio and educational level of the new agricultural business entities are important credit risk indicators. The difficulty of realizing biological assets and whether it is easy to evaluate the value of biological assets have a certain impact on the risk of mortgage loans of biological assets. The number of asset appraisal agencies determines the quality of the regional asset appraisal environment. A good asset appraisal environment can make biological asset valuations more reasonable, thereby reducing the risk of unreasonable biological asset valuations.

4.2. Related Suggestions. Relevant departments of the Chinese government should keep up with the pace of financial innovation and create good conditions for financial institutions to carry out biological asset mortgage loans. First, the “Guarantee Law” and “Property Law” should improve the relevant provisions of biological asset mortgage loans to provide a good legal environment for financial institutions to conduct business. Second, a mechanism for the government, insurance companies, and financial institutions to share the burden should be established. Third, to promote the development of biological asset mortgage loans, government departments should increase financial support to achieve full coverage in terms of policy insurance for farmers, reducing the financial pressure of farmers. Fourth, corresponding guarantee companies can be established to serve Chinese farmers. The biological asset mortgage loans are used as guarantees to reduce the risks borne by financial institutions. At the same time, insurance companies should develop agricultural insurance businesses specifically for biological assets, and for farmers who have purchased policy-based agricultural insurance, insurance costs can be appropriately reduced. This will not only increase the business volume of insurance companies but also bring benefits to farmers. When assets are damaged, insurance companies compensate farmers for their losses in a timely manner so that they have funds to repay financial institutions.

Biological asset valuation is a major problem. Asset valuation industry associations and relevant government departments should help integrate existing biological asset valuation resources, standardize and improve biological asset valuation technology, increase investment, and form a professional biological asset valuation framework. The team conducts targeted assessments of the different growth stages of various assets in the biological asset category, and forms a

unified standard to improve the accuracy of the assessment of their value and continuously reduce the operational risk caused by inaccurate assessment of collateral value, thereby reducing the default risk of lenders and reducing the risk loss of financial institutions. Therefore, in the face of a wide range of mortgage loan needs, we should actively seek cooperation with external valuation agencies, learn from each other, learn from the valuation methods of biological asset valuation institutions, and combine the principle of prudence of financial institutions to rationally value the value of biological assets.

After financial institutions provide loans to rural households, they still need to monitor the biological assets of rural households after the loan and continue to track and analyze them. For animal biological assets, the lender should prevent them from handling or transferring it without authorization; for plant biological assets, attention should be paid to their growth cycle and the impact of natural disasters. Financial institutions should interface with agricultural product information platforms established by relevant departments of the Ministry of Agriculture of various regions, and use remote video monitoring to dynamically grasp the status of biological assets. Monitoring of the mandatory inspection and quarantine of livestock slaughter transactions can effectively prevent the risk of private disposal of live collateral. This “cloud monitoring” model can solve the problem of difficult verification and control of collateral to the greatest extent. In the future, relevant departments will publicize information and update data in a timely manner to achieve data interconnection, reduce information asymmetry, effectively control postloan risks, and continuously improve credit risk prevention and control capabilities.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request (email: 1179188@mail.dhu.edu.cn).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Special fund for Basic Scientific Research of Central University (Grant No: 2232020b-02) and funded by Donghua University.

References

- [1] X. Gao, “Evaluation of the development level of my country’s new agricultural management system,” *Economist (China)*, vol. 9, pp. 73–80, 2017.
- [2] X. Chen and W. Zhu, “Research on the development of new agricultural business entities under the agricultural supply-side reform—based on the investigation and analysis of Sichuan,” *Rural Economy (China)*, vol. 8, pp. 108–114, 2017.

- [3] X. Hu, "Research on the cultivation of new agricultural management subjects under the background of rural revitalization," *Agriculture of Henan (China)*, vol. 26, p. 50, 2019.
- [4] H. Ying, "Research on the dilemma and countermeasures of the industrialization development of the new type of agricultural management," *Agricultural Economy (China)*, vol. 3, pp. 62–64, 2019.
- [5] H. Huang, "Thoughts and suggestions on financial support for the development of new agricultural business entities to help alleviate poverty," *Wuhan Finance (China)*, vol. 11, pp. 63–65, 2018.
- [6] J. Min, "Analysis on the innovation of the new business entity business model—based on the case analysis of Heilongjiang Renfa Cooperative," *Issues in Agricultural Economy (China)*, vol. 10, pp. 50–59, 2018.
- [7] L. J. Mester, "What's the point of credit scoring?" *Business Review*, vol. 3, pp. 3–16, 1997.
- [8] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.
- [9] W. Dobbie, G.-P. Paul, N. Mahoney, and J. Song, "Bad credit, no problem? Credit and labor market consequences of bad credit reports," *Asia-Pacific Journal of Financial Studies*, vol. 49, no. 3, pp. 368–409, 2020.
- [10] C. H. Sossou, F. Noma, and J. A. Yabi, "Rural credit and farms efficiency: modelling farmers credit allocation decisions, evidences from Benin," *Economics Research International*, vol. 2014, Article ID 309352, 2014.
- [11] K. K. Bolarinwa and E. O. Fakoya, "Impact of farm credit on farmers socio-economic status in ogun state Nigeria," *Journal of Social Sciences*, vol. 26, no. 1, pp. 67–71, 2011.
- [12] J. Du and D. Lu, "Operation mode, existing problems and countermeasures of peasant household biological asset mortgage," *Journal Of South China Agricultural University (Social Science Edition) (China)*, vol. 12, no. 2, 2013.
- [13] W. Li, "Exploring and research on the financing model of live animal mortgage—taking hainan prefecture, qinghai province as an example," *Finance Economy (China)*, vol. 20, pp. 161–162, 2014.
- [14] F. Heywood, "Secured transactions: the power of collateral," *Finance and Development*, vol. 6, pp. 44–46, 1996.
- [15] J. Du, D. Lu, and F. Chunyan, "Influencing factors and hierarchical analysis of farmers' biological asset mortgage financing willingness," *Journal of China Agricultural University (China)*, vol. 19, no. 2, pp. 238–244, 2014.
- [16] P. Luo, "Analysis of influencing factors of farmers' biological asset mortgage financing willingness—based on logistic model estimation," *Forestry Economics (China)*, vol. 37, no. 12, pp. 24–29, 2015.
- [17] W. Fan, X. Zhang, and H. Hu, "Study on the selection scheme of biological asset value evaluation method," *Friends of Accounting (China)*, vol. 20, pp. 17–19, 2013.
- [18] E. Baltensperger, "Credit rationing issues and questions," *Journal of Money, Credit and Banking*, vol. 10, no. 2, pp. 170–183, 1978.
- [19] C. John, "A note on the comparison of logit and discriminant models of consumer credit behavior," *Journal of Financial and Quantitative Analysis*, vol. 15, no. 3, 1980.
- [20] W. Liang and S. Wen, "Research on loan default risk assessment of agricultural small and medium-sized enterprises—Based on the data of agriculture, forestry, animal husbandry and fishery enterprises in the "new third board"," *Rural Economy (China)*, vol. 11, pp. 93–100, 2019.
- [21] A. Samad, "Credit risk determinants of bank failure: evidence from US bank failure," *International Business Research*, vol. 5, no. 9, pp. 10–21, 2012.
- [22] M.-F. Hsieh and C.-C. Lee, "Bank liquidity creation, regulations, and credit risk," *The Journal of Finance*, vol. 75, no. 5, pp. 2377–2419, 2020.
- [23] E. Kamaloo and M. S. Abadeh, "Credit risk prediction using fuzzy immune learning," *Advances in Fuzzy Systems*, vol. 2014, Article ID 651324, 2014.
- [24] Y. Yan and W. Fan, "Research on the risk of mortgage financing of live animals," *Rural Economy and Science-Technology (China)*, vol. 21, pp. 163–165, 2016.
- [25] G. Teles, J. Joel, P. C. Rodrigues, R. A. L. Rabêlo, and S. A. Kozlov, "Comparative study of support vector machines and random forests machine learning algorithms on credit operation," *Software: Practice and Experience, Early View*, vol. 21, 2020.
- [26] J. C. Westland, "Predicting credit card fraud with Sarbanes-Oxley assessments and Fama-French risk factors," *Intelligent Systems in Accounting, Finance and Management*, vol. 27, no. 3, pp. 95–107, 2020.
- [27] J. Futoma, J. Morris, and J. Lucas, "A comparison of models for predicting early hospital readmissions," *Journal of Biomedical Informatics*, vol. 56, pp. 229–238, 2015.
- [28] Y. Wu, L. Feng, Z. Jiang, and W. Kong, "Research on risk early warning of P2P online loan platform based on machine learning model," *Journal of Finance and Economics (China)*, vol. 9, pp. 18–25, 2019.
- [29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [30] M. Z. Joharestani, C. Cao, X. Ni et al., "PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data," *Atmosphere*, vol. 10, no. 7, p. 373, 2019.
- [31] Y. Zhou, T. Li, J. Shi, and Z. Qian, "A CEEMDAN and XGBOOST-based approach to forecast crude oil prices," *Complexity*, vol. 2019, pp. 1–15, 2019.
- [32] L. Xia, Y. Zhang, Q. Lu, and G. Tang, "Credit rating method combining XGBoost algorithm and Logistic regression," *Credit Reference (China)*, vol. 37, no. 11, pp. 56–59, 2019.
- [33] C. Li, X. Zheng, Z. Yang, and Li Kuang, "Predicting short-term electricity demand by combining the advantages of ARMA and XGBoost in fog computing environment," *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–18, 2018.