Received April 20, 2019, accepted May 13, 2019, date of publication May 27, 2019, date of current version June 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919256

Data Driven Spatio-Info Network Modeling and Evolution With Population and Economy

JIAN DONG[1], BIN CHEN[1], CHUAN AI[1], PENGFEI ZHANG[1], XIAOGANG QIU[1], AND LINGNAN HE[2]

[1]College of Systems Engineering, National University of Defense Technology, Changsha 410073, China
[2]School of Communication and Design, Sun Yat-sen University, Guangzhou 510000, China

Corresponding authors: Bin Chen (nudtcb9372@gmail.com) and Lingnan He (lingnan.he@qq.com)

This work was supported in part by the National Key Research and Development (R&D) Plan under Grant 2017YFC1200300, in part by the National Natural Science Foundation of China under Grant 71673292 and Grant 71673294, in part by the National Social Science Foundation of China under Grant 17CGL047, and in part by the Guangdong Key Laboratory for Big Data Analysis and Simulation of Public Opinion.

ABSTRACT Spatial interaction is the process that individuals interact with each other at different geographical locations. It attracts much research interests for the increasing data and applications related to spatial interaction. In this paper, a method is proposed to construct the spatio-info network with the dataset from WeChat. The correlation between human factors and statistics characteristics of the network is analyzed and confirmed, and then, the gross domestic product (GDP) and demographics are integrated into gravity model to model the spatio-info network. The likelihood method is used to solving the parameters and evaluates the four models; it is found that the GDP-GDP-distance (GGD) and population-population-distance (PPD) are similar and much better than the other two models. Finally, topological characteristics and community structure of the evolution network are analyzed to evaluate the models. It is found that evolution networks of the two models are almost consistent to origin network, and PPD models are better. It is concluded that the gravity model and human factors can be used to model the spatio-info network. This paper can be used to predict the communication amount of different regions in online social media dynamically. Naturally, this will help the mobile communication infrastructure construction, especially for a new generation of technology, such as 5G, or for regions with poor infrastructure. In addition, it will also help the software service providers configure server and advertising resources.

INDEX TERMS Spatio-info network, law of universal gravitation, GDP, demographics.

I. INTRODUCTION

There are different types of information dissemination methods in Internet, including Web, FTP, E-mail and Telnet. In recent years, online social networks have spread rapidly around the world and are now one of the most popular types of websites on the Internet [1]. The popularity of the website provides a lot of opportunities to study the characteristics of online social networks [2]. Research on social networks involves many topics, such as the basic features of social networks ([2]–[5]), user behavior characteristics ([6]–[8]) online propagation ([9], [10]), and the evolution of online social networks ([11]) and so on.

Spatial interaction is the process that an entity makes contact, demand or supply decisions, and location choices at different geographical locations [12] For example, trade between different countries or regions, human migration between cities or countries, and communication between people in different cities through telephone or online social media. In online social network, spatial interaction is formed by users at different spatial locations, to send, view and forward information. Naturally, spatial interactions can be described with complex network methods [13]. Nodes represent spatial locations, which can be cities, provinces, or countries. Edges represent interactions of entities in different spatial locations. Studying the characteristics of spatial interaction networks in social networks and their evolutionary mechanisms is of great significance for providing location-based business services, planning and managing communication network facilities, and formulating regional economic development policies.

The spatial interaction characteristics of online social networks have attracted wide attention. From the perspective of spatial information diffusion, geoscience research based on

The associate editor coordinating the review of this manuscript and approving it for publication was Yin Zhang.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

mobile communication data has two important results. The first is that the law of universal gravitation can be used to fit the density of interactions between cities' centers [14], [15]. The other is about community detection, some studies have found that the detected communities match the administrative divisions [16], [17] very well. Moyano *et al.* [18] and others found that strong interactions usually indicate that the economic relationship between the two sides is strong. These studies follow the tradition that spatial interactions can represent the theory of geographic features, which was established by Ullman and Boyce [19] and persisted and developed by Noronha and Goodchild [20] and Castells [21].

This paper studies the evolution mechanism of cities interactive networks formed by the information dissemination process on online social networks. The nodes of this network are cities, and the interactions among cities of users are regarded as edges. At present, many studies on the evolution of social networks are considered in the perspective of nodes by adding nodes to the network with given rules and establishing links with nodes in the network [22], [23]. However, this evolutionary mechanism cannot be applied to cities interactive networks. In the process of information propagation among cities, the geographic locations is fixed, and the interaction between cities changed continuously. Therefore, we consider the evolution model of cities online interactive network from the perspective of the edges. In the model, the nodes are fixed, and the evolution process consists of two parts: the link arrival time and the preferential attachment. We consider the network as a sequence of links, the generation of each link consists of two aspects: one is when a link happens in the networks, and another is that in which cities a user sends the information and another user receive it.

The remainder of the paper is organized as follows. In section two the related works are introduced, including spatio-info networks, the centrality and community of network, the extension of gravity law and the method of model evaluation by likelihood. Section three introduces data and spatio-info networks, the generation method of Spatio-info Network is proposed. The characteristics of the network centrality and its relationship with population and economy are analyzed in section four. Section five proposed the model, and Section six generate an evolutionary network with the model and compares the evolutionary network with real data. Section seven discussed the results with related work and the final section summarizes the work of the paper.

II. RELATED WORK

This paper first constructs Spatio-network taking cities as a node, which is similar to the research method of Blondel *et al.* [16]. They analyze the data set containing 200 million mobile phone communication records and construct a network with cities as the nodes, and adopt the method of community detection to divide network nodes into clusters to research the existence of regions and boundaries.

There has been a lot of researches on spatial information interaction. Research related to community detection based on mobile communication data mainly involves four aspects, in addition to the above-mentioned city-based networks, there are individuals based [24]–[26], pixel grid [17] based, and Tyson polygons based [27], [28] centered on the mobile phone base station. The different types of networks are determined by different data sets. The nodes of these networks are geocoded into the cities, square pixels and cells based on consumption data or the geographic location of the caller. This paper summarizes the methods of these researches and proposes a method for Spatio-info networks generation based on the information interaction network among individuals.

A. NETWORK CENTRALITY

The nodes' degree measures the local influence of the nodes in the network. The betweenness centrality of the nodes measures the nodes' influence in the global context. The betweenness centrality of a specific node is the ratio of the shortest paths between any two pairs of nodes that pass the node [29].

$$C_B(k) = \sum\nolimits_{i \neq j = k}^{N} \frac{\sigma_{ij}(k)}{\sigma_{ij}} \tag{1}$$

As shown in eq. (1), Where σ_{ij} is the number of the shortest path between nodes i to j, and, $\sigma_{ij}(k)$ is the number of the shortest path that pass node K. The degree and betweenness centrality of a node measure different aspects of the node's position in the network.

Barrat *et al.* [3] found the power-law distribution with the increase of node degree in the betweenness centrality. Guimera and Amaral [30] found that nodes with higher degrees do not always have high betweenness centrality. Onnela *et al.* [31] found that in mobile phone communication networks, the betweenness centrality of the edge is inversely related to the weight of the edge. The analysis of the betweenness centrality of these studies is based solely on individual networks. In the work of Guanghua *et al.* [28], the concept of geographical center is analyzed from the perspective of geographic information with the method of .betweenness centrality analysis and community detection. Based on the GDP and demographic data of each city, this paper further analyzes and discovers the high correlation between economic, demographic centers and network centers.

B. GRAVITY LAW

Researchers found there is a phenomenon similar to Law of Universal Gravity, in situations such as crowd travel networks, population migration networks, and commodity trade networks. Then gravity model for the spatial interaction is proposed by analogy with Newton's Law of Universal Gravity. The gravity model provides an estimate of the flow between two or more regions (e.g., the number of trips by residents, the number of trades in goods). In a spatial interaction network, the gravity model can be interpreted as that, the frequency of interaction between two nodes is proportional to the strength of the two nodes (demographics, economy), and inversely proportional to the distance of the two nodes. The gravity model naturally becomes a classic model for interpreting and predicting the interaction strength of spatial networks. It is widely used in transportation planning, immigration, international trade [31], [32] and disease transmission [34] and many other fields.

Although the gravity model is simple, intuitive and easy for calculation, and incorporates geographic factors, it lacks a rigorous theoretical foundation, and these studies are modeled from a static perspective.

Therefore, the main content of this paper is to construct the evolutionary network model of Spatio-info networks by integrating economic and demographic factors in the context of gravity law. Its main contribution is to analyze the correlation of economic and demographic factors with network centers, provide a more rigorous theoretical basis for the gravitational model, and constructing models from the perspective of evolutionary networks rather than static networks. It provides strong theoretical support for network prediction research and the Internet management.

In order to illustrate the advantages of the model, a comparison method that compares the maximum likelihood values is introduced. The likelihood is usually used to numerically compare a series of models and to select the best model (and parameters) to interpret the data according to the maximum likelihood criterion [35]. As our understanding of real-world networks improves, likelihood remains unchanged while the generative models improve to incorporate the new understanding. Success in modeling can therefore be effectively tracked [22]. Likelihood have been commonly used in estimation of network model parameters [36], [37] and model selection [22], [38]. Therefore, we use the model likelihood method to evaluate and compare different network evolution models with empirical data.

III. NETWORK DEFINITION AND CONSTRUCTION

WeChat is the most widely and frequently used online social media software for Chinese, supporting the PC and mobile termination. WeChat users will leave a record when they click to browse the pages shared by others, and this is our data source. The historical records of a large number of pages transmitted between users are collected by the company (Fabonacci), forming a data set.

Each record in the data refers to the behavior of a user's one click and browsing the webpage, including the sharer's ID, the viewer's ID, the page ID, the viewer's IP, and the browsing time. To protect the user's privacy, all user IDs are unique to the user's WeChat ID, but they are not the same, so it is impossible to analyze the data for a specific real-life user. Through the query of the IP library, the geographical location of the users are obtained. Jian *et al.* [39] analyzed several major domestic and foreign IP address libraries, including *IP2Location Lite*, *GeoLite2* abroad, domestic *Pure IP Address Library*, *Taobao IP Address Library*, *Sina IP Address Library* and*Baidu IP Address Library*. They compared coverage and coincidence of these address libraries, and concluded that the *Taobao IP Address Library* has the highest credibility and credibility at the city level. Since the IP address library is difficult to be corrected, and the lack of information about the establishment methods of the foreign IP address libraries, the reliability of the *IP2Location Lite* and *GeoLite2 IP address pools* cannot be known. In this paper, *Taobao IP Address Library* is used to geographically locate the IP addresses involved in the data.

A. NOTATION

There are many ways to represent a network. In this paper, M is used to indicate the number of nodes in the network, Z is the network of M*M, $\mathbb{N}$ is used to represent the connected space, and N is the number of edges. z_{sr} represents the connection weight between node s and node r, and non-zero z_{sr} indicates that there is one link at least between node s and node r. In this paper, the Spatio-info Network refers to the spatial information interaction network that regards cities as nodes. It is transferred from the online social network in which users are nodes according to the user's IP. The specific construction method is elaborated in section three.

B. THE CONSTRUCTION OF SPATIO-INFO NETWORK

A total of 277 cities have appeared in the data. There is no affiliation among these cities. Their administrative areas on the map can almost form a complete map of China except Taiwan, there are some cities that we did not include because we cannot get the GDP or demographic data. M^h is used to indicate the number of users in the data set, and M^c indicates the number of cities in the data set. First of all, the interactive records of data can be considered as the sequence of links between users from the perspective of information dissemination, and the amount of the sequence is represented by N. $\Omega(A)$ is an operator, and the value is 1 if A is true. The users' network can be represented as $z_{ij}^{(N)}$ as shown in formula (2).

$$\begin{aligned} y^{(n)} &= (s_n, r_n)\; n = 1, \dots, N \\ Y &= \left\{y^{(1)}, y^{(2)}, \dots, y^{(N)}\right\} \\ z_{ij}^{(N)} &= \sum_{n=1}^{N} \Omega\left(s_n = i, r_n = j\right), \quad i,\, j \in \left\{1, \dots M^h\right\} \end{aligned} \tag{2}$$

Each user node i has a city assignment g_i, and the interaction between the users is regarded as the interaction between the cities. Each interaction record can be regarded as a directed link, which is expressed as a city pair. The data set can be seen as a sequence of city pairs. The Spatio-info Network among cities can also be represented in the form of adjacency matrix $z_{mk}^{,(N)}$ as shown in formula (3).

$$\begin{aligned} x^{(n)} &= \left(g_{s_n}, g_{r_n}\right) \\ X &= \left\{x^{(1)}, x^{(2)}, \dots, x^{(N)}\right\} \\ z_{mk}^{,(N)} &= \sum_{n=1}^{N} \Omega\left(g_{s_n} = m, g_{r_n} = k\right), \quad m,\, k \in \left\{1, \dots M^c\right\} \end{aligned} \tag{3}$$

The spatio-info network Z_T is constructed from the real data set during the period from 00:00 on July 2, 2016, to 24:00

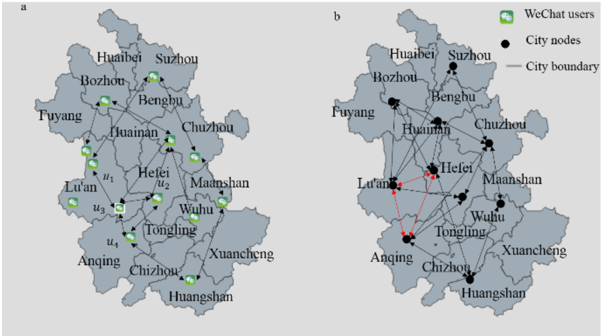


FIGURE 1. Schematic diagram of network generation method.

TABLE 1. Basic attributes of a city interactive network Z_T.

Symbol	Explanations	Value
T	Time period	7/2/2016-7/8/2016
M	Number of nodes	277
M_T	Number of links	514177
k_T^{avg}	Average degree	77.491
M_T^{se}	Number of self-connected links	275
ρ_T	Density	0.281
L_T	Clustering coefficient	0.509

on July 8, 2016. Table 1 lists the basic characteristics of the network.

According to the basic characteristics of the network listed in table 2, an overall understanding of the interaction between web pages in WeChat is presented. The network involves 277 nodes and 514177 links. On average, each node has links only with 77.491 different nodes, and the density of the network is only 0.281. It can be seen that although WeChat has a large number of users in China, covering all cities in the country, in a period of 7days, there are city pairs that do not have information exchange in this platform. The average shortest path length of the network is 1.727. That is to say, from one node to another, it passes through 1.727 links on average. The average clustering coefficient of the network is 0.509. The network has obvious characteristics of small world. There are 275 nodes with self- connecting links in the network. There are 514,177 interactive records in the data, of which there are 312,292 internal interactions, accounting for 60.7%. It can be seen that users are more inclined to interact with users in the same city.

C. CENTRALITY OF THE SPATIO-INFO NETWORK

The degree and betweenness centrality of the resulting network (Spatio-info network) are calculated, and compare the rank of the degree and betweenness centrality are compared with the rank of city nodes' GDP and population. As shown in figure 2 below, it can be seen that GDP and degree and betweenness centrality are significantly statistically correlated with demographics and GDP. The correlation coefficient between demographics and degree is 0.74. The

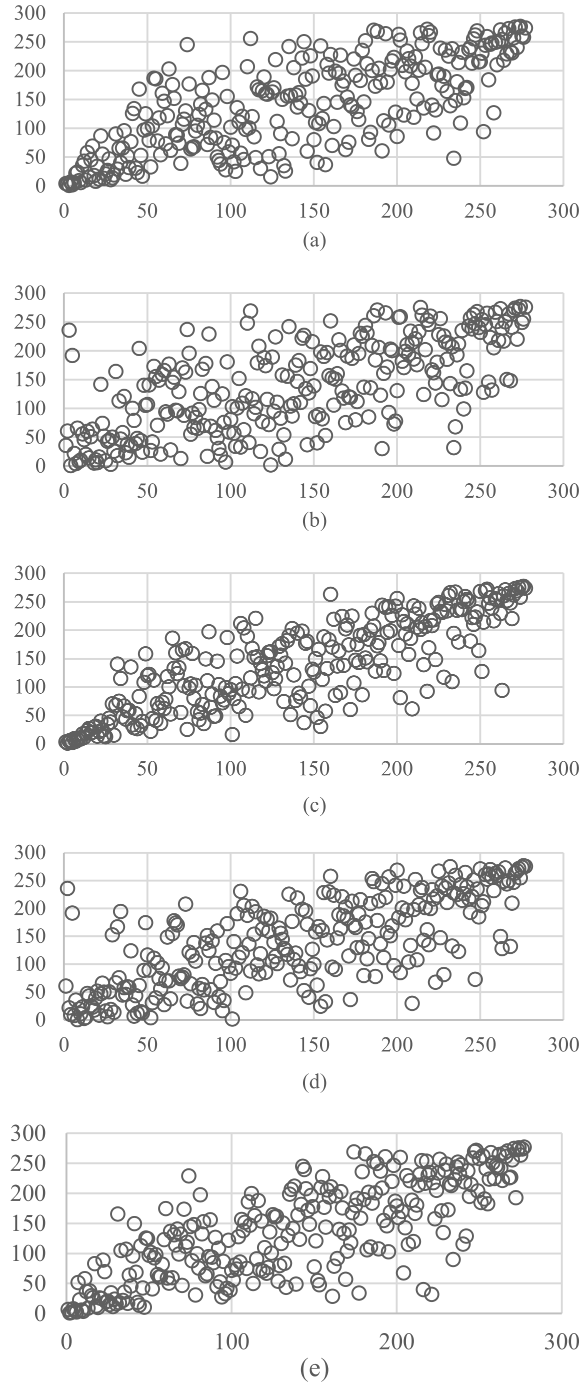


FIGURE 2. The corelation between the factors of the network. (a) $R^2 = 0.74$, $P < 0.0001$ (population-degree). (b) $R^2 = 0.68$, $P < 0.0001$ (population-betweenness). (c) $R^2 = 0.82$, $P < 0.0001$ (GDP-degree). (d) $R^2 = 0.73$, $P < 0.0001$ (GDP-betweenness). (e) $R^2 = 0.78$, $P < 0.0001$ (GDP-population).

correlation coefficient between demographics and betweenness centrality is 0.68. The correlation coefficient between GDP and degree is 0.82. The correlation coefficient between GDP and betweenness centrality is 0.73. It can be found that the correlation coefficient between GDP and network centrality (degree and median centrality) is greater than that between population and network centrality (degree and median centrality).

In complex network theory, the degree and betweenness centrality of nodes can reflect the importance of nodes in the network. The degree represents the local centrality and importance of the nodes. The betweenness centrality reflects the centrality and importance of the node in the network in global perspective. Cities with highest GDP are usually the economic centers in a region (such as a province), and cities with larger demographics are usually population and political centers. The demographics, GDP are significantly correlated with the degree and betweenness centrality of city nodes in the network. It can be said that the local and global centers of the Spatio-info Network are consistent with the political and economic population centers. From this perspective, the network can be modeled with factors of demographics and GDP. Modeling and examination can also in turn describe the impact of demographics and GDP in the Spatio-info network. Therefore, it can be assumed that for a new link the probability to choose one city as a sender or receiver's location assignment is positively correlated with that city's demographic or GDP.

According to the theory of gravity model, it can be further assumed that the probability to choose one city as a sender or receiver's location assignment is negatively correlated with the distance between two cities.

At the same time, as can be seen from figure 2(e), the correlation between economic and demographic data is very strong, and the correlation of economy or demographic with network center is nearly the same. Naturally, the economy, demographic and distance should be simultaneously taken into consideration. However, because the correlation coefficient between GDP and demographic ranking is also significantly correlated, it is not meaningful to consider GDP and demographic together. From the perspective of model complexity, it will increase the parameters and brings a lot of computation for parameter solving and the process of getting an evolutionary network. Therefore, two different models will be considered in this paper. One model considers GDP and distance, and another one considers population and Distance. The two models with economic-only models and demographic -only models will be compared to find the right modeling method.

IV. PREFERENTIAL ATTACHMENT FOR LINKS

This article considers the evolution of the network from the perspective of constantly adding new links. When a link joins the network, the behavior of one user clicking or sharing another user's webpage in the network is simulated. The links source node and destination node should be sampled according to a certain mechanism. From the content of the third section, the economy, population and distance have a significant impact on the links. Therefore, a priority connection mechanism is built based on four different models. The four models considers different factors combination: the economy (GDP-GDP) only, the demographics. (PP) only, economy and distance (GGD), and demographics and distance (PPD).

GG: As shown in eq. (4), it is assumed that the probability of choose the source node and the destination node is positively correlated with the GDP of the city node for a link that arrived at t. The two parameter are α and β:

$$P_{GG}(e^t_{u,v}) = \frac{[k_t(v)]^\alpha [k_t(u)]^\beta}{\sum\limits_{i,j\in V} [k_t(i)]^\alpha [k_t(j)]^\beta} \tag{4}$$

GGD: As shown in eq. (5), it is assumed that the probability of choose the source node and the destination node is positively correlated with the GDP and negatively correlated with distance of the two city nodes for a link that arrived at t. The three parameter are α, β for nodes and γ for distance:

$$P_{GGD}(e^t_{u,v}) = \frac{[k_t(v)]^\alpha [k_t(u)]^\beta d(u,v)^\gamma}{\sum\limits_{i,j\in V} [k_t(i)]^\alpha [k_t(j)]^\beta d(i,j)^\gamma} \tag{5}$$

PP: As shown in eq. (6), it is assumed that the probability of choose the source node and the destination node is positively correlated with the demographic of the city node for a link that arrived at t. The two parameter are α and β:

$$P_{PP}(e^t_{u,v}) = \frac{[k_t(v)]^\alpha [k_t(u)]^\beta}{\sum\limits_{i,j\in V} [k_t(i)]^\alpha [k_t(j)]^\beta} \tag{6}$$

PPD: As shown in eq. (7), it is assumed that the probability of choose the source node and the destination node is positively correlated with the demographic and negatively correlated with distance of the two city nodes for a link that arrived at t. The three parameter are α, β for nodes and γ for distance:

$$P_{PPD}(e^t_{u,v}) = \frac{[k_t(v)]^\alpha [k_t(u)]^\beta d(u,v)^\gamma}{\sum\limits_{i,j\in V} [k_t(i)]^\alpha [k_t(j)]^\beta d(i,j)^\gamma} \tag{7}$$

The method of maximum likelihood estimation is user to estimate the parameters of different models. Given the real data of network evolution, we can test the model performance by likelihood. The likelihood method is usually used to numerically compare a series of models and to select the best model (and parameters) to interpret the data based on the maximum likelihood criterion. For the model M, the probability that a link $e^t_{u,v}$ takes u as the source node, and v as the destination node $P_M(e^t_{u,v})$ should be calculated. It should reach the maximum in the real data. And through model M, the probability that network Z_T is generated in time T is $P_M(Z_T)$. It also should reach the maximum with the real condition, which is shown in real data.

The log-likelihood value is shown in eq. (8):

$$\log(P_M(Z_T)) = \log(\prod_{t\in T} P_M(e^t_{u,v})) = \sum_t \log(P_M(e^t_{u,v})) \tag{8}$$

Since the Spatio-info network has self-connected links. It is assumed that if a link is self-connected, the distance of two position is 20 kilometers, although the distance of a same city should be the same (each city is considered

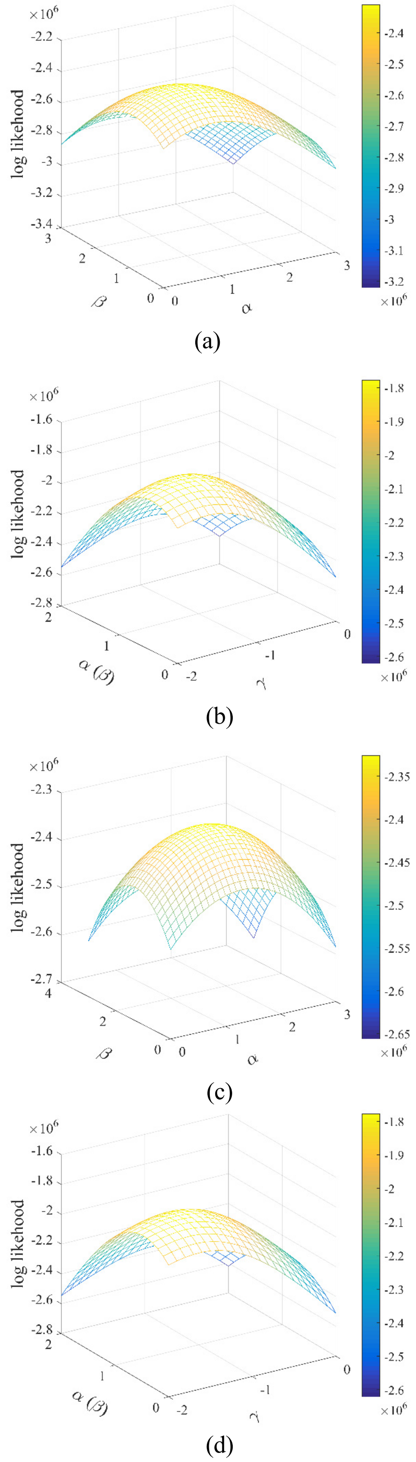


FIGURE 3. (a) The relationship between the maximum log-likelihood of the GG model and the parameter α and. β. (b) The relationship between the maximum log-likelihood of the GGD model and the parameter α, β and γ. (c) The relationship between the maximum log-likelihood of the PP model and the parameter α and. β. (d) The relationship between the log maximum likelihood of the PPD model and the parameter α, β and γ.

a circle, 20 kilometers is assumed the radius of all cities averagely). Figure 3 shows the relationship between the log-maximum likelihood and the parameters of different models. It can be seen that the log likelihood of the four models is a convex function of the model parameters, and the maximum likelihood of each model can be found to estimate the optimal parameters. Among them, the PP model obtains the maximum likelihood value when $\alpha = 1.3$ and $\beta = 1.3$. The GG model obtains the maximum likelihood value at $\alpha = 0.9$ and $\beta = 0.9$. The PPD model is $\alpha = \beta = 0.9$, $\gamma = 1.6$, when the maximum likelihood value is obtained, the GGD model obtains the maximum likelihood value when $\alpha = \beta = 0.5$ and $\gamma = 1.6$.

TABLE 2. Model parameters and maximum log-likelihood.

Model	Parameters	Value
PP	α=1.3, β=1.3	-2326048
PPD	α=β=0.9, γ=1.6	-1776805
GG	α=0.9, β=0.9	-2310433
GGD	α=β=0.5, γ=1.6	-1778219

Table 2 lists the maximum log-likelihood values and the corresponding parameters for different models. It can be seen from table 2 and the figure 3 that the maximum log-likelihood value obtained by the PP model is lower than that of the GG model, and that value obtained by the GGD model is lower than that of the PPD model. The likelihood value of the model exhibits the following relationship as in formula (9):

$$\log_{PP} < \log_{GG} < \log_{GGD} < \log_{PPD} \tag{9}$$

Obviously, the likelihood values obtained by the GGD and PPD models are relatively large. It shows that these two models PPD and GGD are more suitable for modeling the Spatio-info network than the other two models. Compared with the GG and PP model, the maximum log-likelihood values of GPD and PPD increase by about 23%. And PPD model is slightly better.

V. EXPERIMENTS

In order to further study the evolution performance of PPD and GGD models, we conduct some experiments based on network of real data.

First, a real network Z_T^R based on 7-day real data is taken into consideration. Based on the origin network, the optimal parameters of PPD and GGD model can be obtained.

With these parameters, two evolution networks of seven-day's period can be produced based on the PPD and GGD model respectively. At beginning, when t = 0, there is no link in each network. And from t = 0 to t = T, links are added into the networks one by one with the methods and parameters of PPD and GGD.

Finally, the evolution networks Z_T^P and Z_T^G of seven days are obtained. The topology characteristics and community structure of evolution networks are analyzed by comparing with the real network Z_T^R.

A. CENTRALITY OF THE SPATIO-INFO NETWORK

Figure 4 shows the comparison of the topological features of the evolved network with the original network. Red data

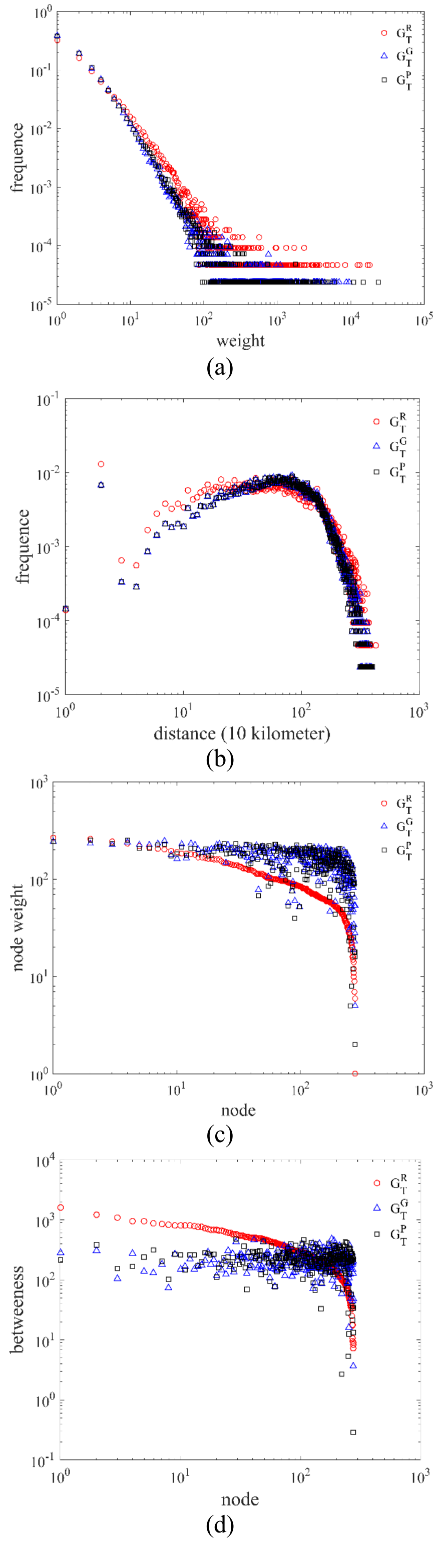


FIGURE 4. Topological characteristics of evolution networks and origin network. (a) The edges' weight distribution. (b) The distance distribution (c) The nodes' weight distribution. (d) The betweenness centrality distribution.

points represent features of Z_T^R, data points of blue triangles and blue squares represent features of evolutionary network Z_T^G and Z_T^P.

Figure 4(a) shows the distribution of the edges' weight. The weight of an edge represents the number of links between two cities. It can be seen that the edges' weight distribution of the real network and the evolution network are nearly the same and obey the power-law distribution.

Figure 4(b) shows the distance distribution of any two city nodes. This is an attribute that reflects the characteristics of geospatial. It reaches the biggest around 100 kilometers. As the distance increases, the probability of links generation is smaller. In addition, 20 kilometers is also the high-frequency distance of links (the distance of interaction within the city is recorded as 20 kilometers), indicating that the links within the city occupies a large proportion.

Figures 4(c) and (d) take the node properties into consideration. Figure 4 (c) shows the weight distribution of the nodes. The abscissa is the node number, and the number order is arranged in descending order of the node point weight. The weight of a node is the sum of the weights of all connected edges connected to the node, reflecting the amount of interaction between the node and other nodes. The weight of a node is the sum of the weights of all edges connected to the node, reflecting the amount of interaction that the city generated with other cities.

Figure 4(d) shows the betweenness centrality of the nodes. The abscissa is the node number, and the number order is sorted in descending order of the betweenness centrality. Betweenness centrality measures the role of a node in global view. It can be found that there are obvious differences in the nodes' weight and the betweenness centrality between the evolutionary network and origin network, but the overall trend is consistent.

At the same time, from the purpose of comparing the two models of PPD and GGD, there are not significant difference between the two evolutionary networks obtained by the two models. Especially, the nodes' weight distribution and distance distribution are basically consistent and very close to the real network. The similarity of the two evolutionary networks can be explained by the same model structure and the significant correlation between the city's GDP and the demographics.

B. COMMUNITY STRUCTURE COMPARISON

Community detection is an important research method in the research of Spatio-info networks. The study by Guanghua *et al.* [28] and Ai *et al.* [40] found that such networks have important characteristics in the community structure. In the results obtained by the community detection, the nodes (city or base station) belonging to the same community are geographically close, and the boundaries of the community geographically agree well with the boundaries of administrative divisions. Therefore, the community structure is also an important nature of this type of network. The community detection method is used to analyze the original network and the evolution network. The results are compared to verify that the network evolved from the model has similar community structure characteristics of the original network.

TABLE 3. The community detection results of three networks.

Detection results	Modularity	Community numbers
$C(Z_T^R)$	0.344	22
$C(Z_T^G)$	0.482	24
$C(Z_T^P)$	0.442	25

TABLE 4. The similarity of different divisions.

	$C(Z_T^R)$	$C(Z_T^G)$	$C(Z_T^P)$	$C(ad)$
$C(Z_T^R)$	1	0.68	0.63	0.88
$C(Z_T^G)$	0.68	1	0.95	0.66
$C(Z_T^P)$	0.63	0.95	1	0.72
$C(ad)$	0.88	0.66	0.72	1

For a specific network Z, node $s \in [1, M^c]$, in a kind of nodes' division, c_s is used to represent the community assignment, $C(Z)$ represents the list of all nodes' community assignments as shown in (9).

$$C(Z) = \{c_1, c_2, \ldots, c_{M^c}\} \tag{10}$$

Therefore, the community detection results of network Z_T^R, Z_T^G, Z_T^P is $C(Z_T^R)$, $C(Z_T^G)$, $C(Z_T^P)$. The administrative division of the cities in province-level region is $C\ (ad)$.

As shown in table 3, Fast_unfolding [41] is used to analyze the networks Z_T^R, Z_T^G, Z_T^P. There are 22, 24, 25 communities and modularity is 0.344, 0.482, 0.442 in result of $C(Z_T^R)$, $C(Z_T^G)$, $C(Z_T^P)$ respectively.

The method of community detection divide the nodes into different communities. In order to further study whether the community detection results are consistent for different networks, the Adjusted Rand Index (ARI) is used to evaluate the similarity between two results. The interval of ARI is [0, 1]. A ARI index that is closed to zero, it indicates that two divisions have big difference, and two divisions is nearly the same when it is close to 1.

In the work of Guanghua *et al.* [28] and Ai *et al.* [40] found that the structure of Spatio-info networks was highly consistent with administrative division. Therefore, the community detection results of origin network and evolution network are compared in this paper. As shown in table 4, the ARI represents the similarity of two corresponding divisions.

This table is a matrix, it can be found that, the ARI between $C(Z_T^G)$, $C(Z_T^R)$ is 0.68, ARI between $C(Z_T^P)$, $C(Z_T^R)$ is 0.63, ARI between $C\ (ad)$, $C(Z_T^R)$ is 0.88. It can be found that the modularity and community numbers of evolution network and origin network of real data have little difference. And for the index of ARI, the result of PPD and GGD are nearly the same. They are all similar to the results of origin network, but the ARI of PPD model's network with origin network is higher than the other one. It indicates that PPD model is slightly better which is also shown in formulation 8.

VI. DISCUSSION AND CONCLUSION

The method of constructing Spatio-info network based on actual data is introduced in this paper. This kind of networks has characteristics of social networks such as large scale, complex structure and scale-free. They are also coupled with complex geographic information. This kind of networks are very important for studying the geographical characteristics of information dissemination. At present, there are many different construction methods for such networks, and they also reflect different granularities in the geographical scope. For example, one node in this paper represents a city's regional scope, and in some research nodes are base station radiation ranges for mobile termination. Some other research constructs a network with geographical blocks in pixel grids as nodes. The division method is worthy of study. For example, according to the administrative division of the city, the distribution of the base station on the ground is based on the division of the Thiessen polygon, or uniform matrix. Different division methods are mainly chosen based on data characteristics or research purposes. The links are obtained from the records of movement of population, communication in Internet or mobile communication with cell phones. The method proposed in this paper summarizes these different network construction methods and makes them in a unified and normative form.

Secondly, this paper studies the relationship between two kinds of factors, which are human factors and statistical features of Spatio-info network. Specifically, GDP and demographics are human factors, these factors reflect the economy and population condition in the region of a city. And statistical features of Spatio-info network are degree and betweenness centrality for they reflect the city nodes' important role in local and global respectively. The work in this paper shows that the factors of demographics and GDP are significantly related to the characteristics of degree and betweenness centrality. From this point, it can be proved that the local center and the global center in the Spatio-info network are in good agreement with the population-concentrated and economically developed areas. However, we believe that there are more factors that have strong correlation with such kind of networks, for example, population of different ages may have different relationships with the Spatio-info network, and whether Internet companies have influence on the Spatio-info network is worth studying.

Due to the relationship between two kinds of factors, it is naturally model the Spatio-info network based on demographics and GDP. At the same time, because the network has geographic information, this paper takes the distance between two cities as one of the factors when modeling. The city government's location is considered as the city center precisely. These factors are integrated into gravity models to model the Spatio-info network. Though many factors are considered, there are some factors that may have influence on the Spatio-info network, such as economic structure, culture difference, and language. And as for the distance, taking the distance of two cities' government centers as two cities' distance is simple and easy to calculate. However, in modern society, distance is heavily influenced by transportation. There are many factors have influence on distance. Airplane, high-speed railway saves much time in the trip between two cities, and cities in the same province that have intercity high-speed rail have shorter distances. The distance is hard to quantify.

Gravity model is a generic that is used in many research, especially in the research of Spatio-info network and similar networks. In this paper, an evolutionary network model is constructed. And the process of network evolution can be reproduced.

Finally, GGD and PPD models are used to generate evolutionary network. Evolutionary network is compared with origin network from basic topological features and community structure. In the basic topological characteristics, the evolutionary network is mostly consistent with the origin network. In the community structure, the results of community detection can be found. The similarity between the real network and the evolutionary network generated by the two models is obvious, and the PPD model gets better results.

In the future work, more factors should be included. And more datasets of different platforms or countries can help to construct more detailed and abundant experiment about this model. Finally, a prediction research for real deployment of network services should be constructed to test this model, which will demonstrate the value and significance of the model.

REFERENCES

[1] R. B. Zadeh, A. Goel, K. Munagala, and A. Sharma, "On the precision of social and information networks," in *Proc. 1st ACM Conf. Online Social Netw.*, Oct. 2013, pp. 63–74.

[2] A. Mislove, M. Marcon, K. P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, Oct. 2007, pp. 29–42.

[3] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proc. Nat. Acad. Sci.*, vol. 101, no. 11, pp. 3747–3752, 2004. doi: 10.1073/pnas.0400087101.

[4] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small—World' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.

[5] B. Albert-László and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, Oct. 1999.

[6] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *Proc. 9th ACM SIGCOMM Conf. Internet Meas.*, Chicago, Illinois, USA, Nov. 2009, pp. 49–62.

[7] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger, "Understanding online social network usage from a network perspective," in *Proc. 9th ACM SIGCOMM Conf. Internet Meas.*, Chicago, Illinois, USA, Nov. 2009, pp. 35–48.

[8] A. Nazir, S. Raza, and C.-N. Chuah, "Unveiling Facebook: A measurement study of social network based applications," in *Proc. 8th ACM SIGCOMM Conf. Internet Meas.*, Oct. 2008, pp. 43–56.

[9] M. Granovetter, "Threshold models of collective behavior," *Amer. J. Sociology*, vol. 83, no. 6, pp. 1420–1443, May 1978. doi: 10.1086/226707.

[10] D. J. Watts, "A simple model of global cascades on random networks," *Proc. Nat. Acad. Sci.*, vol. 99, no. 9, pp. 5766–5771, Apr. 2002. doi: 10.1073/pnas.082090499.

[11] N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, and D. Song, "Evolution of social-attribute networks: Measurements, modeling, and implications using Google+," in *Proc. Internet Meas. Conf.*, Nov. 2012, pp. 131–144.

[12] R. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interaction models of images," *IEEE Trans. Inf. Theory*, vol. 29, no. 1, pp. 60–72, Jan. 1983.

[13] B. Dejon, "Attraction-regulated dynamic equilibrium models of migration of the multiplicative type," in *Proc. Evolving Geograph. Struct.*, Jun. 1983, pp. 12–23.

[14] G. Krings, F. Calabrese, C. Ratti, and V. D. Blonde, "Urban gravity: A model for inter-city telecommunication flows," *J. Stat. Mech., Theory Exp.*, vol. 7, Jul. 2009, Art. no. L07003.

[15] C. Roth, S. M. Kang, M. Batty, and M. Barthélemy, "Structure of urban movements: Polycentric activity and entangled hierarchical flows," *PloS one*, vol. 6, no. 1, Jan. 2011, Art. no. e15923. doi: 10.1371/journal.pone.0015923.

[16] V. Blondel, G. Krings, and I. Thomas, "Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone," *Brussels Stud.*, vol. 42, pp. 1–13, Oct. 2010.

[17] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz, "Redrawing the map of Great Britain from a network of human interactions," *PloS one*, vol. 5, no. 12, 2010, Art. no. e14248. doi: 10.1371/journal.pone.0014248.

[18] L. G. Moyano, O. R. M. Thomae, and E. Frias-Martinez, "Uncovering the Spatio-temporal structure of social networks using cell phone records," in *Proc. IEEE 12th Int. Conf. Data Mining Workshops*, Dec. 2012, pp. 242–249.

[19] E. Ullman and R. R. Boyce, *Geography as Spatial Interaction*, vol. 44. Washington, DC, USA: Univ. Washington Press, 1980.

[20] V. T. Noronha, and M. F. Goodchild, "Modeling interregional interaction: Implications for defining functional regions," *Ann. Assoc. Amer. Geographers*, vol. 82, no. 1, pp. 86–102, Mar. 2005.

[21] M. Castells, *Rise of the Network Society: The Information Age: Economy, Society and Culture*. Hoboken, NJ, USA: Blackwell Publishers, 1996.

[22] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2008, pp. 462–470.

[23] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Phys. A, Stat. Mech. Its Appl.*, vol. 311, pp. 590–614, Aug. 2001.

[24] J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. D. Menezes, K. Kaski, A. L. Barabási, and J. Kertész, "Analysis of a large-scale weighted network of one-to-one human communication," *J. Phys.*, vol. 9, no. 6, p. 179. Jun. 2007.

[25] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, pp. 664–667. Apr. 2007.

[26] R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. V. Dooren, "Geographical dispersal of mobile communication networks," *Phys. A, Stat. Mech. Its Appl.*, vol. 387, no. 21, pp. 5317–5325, 2008. doi: 10.1016/j.physa.2008.05.014.

[27] S. Gao, Y. Liu, Y. Wang, and X. Ma, "Discovering spatial interaction communities from mobile phone data," *Trans. GIS*, vol. 17, no. 3, pp. 463–481, Jun. 2013. doi: 10.1111/tgis.12042.

[28] G. Chi, J.-C. Thill, D. Tong, L. Shi, and Y. Liu, "Uncovering regional characteristics from mobile phone data: A network science approach," *Papers Regional Sci.*, vol. 95, no. 3, pp. 613–631, Aug. 2016. doi: 10.1111/pirs.12149.

[29] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, Mar. 1977. doi: 10.2307/3033543.

[30] R. Guimerà and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, pp. 895–900, Feb. 2005.

[31] J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks," *Proc. Nat. Acad. Sci.*, vol. 104, no. 18, pp. 7332–7336, May 2007.

[32] M. Dueñas and G. Fagiolo, "Modeling the international-trade network: A gravity approach," *J. Econ. Interact. Coordination*, vol. 8, no. 1, pp. 155–178, Apr. 2013.

[33] C. Carrère, "Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model," *Eur. Econ. Rev.*, vol. 50, no. 2, pp. 223–247, Feb. 2006.

[34] D.-F. Lin, Z.-k. Qi, and Q.-L. Xia, "Design and identification on parameters of proportional navigation guidance law with gravity over compensation," *J. Syst. Simul.*, vol. 10, no. 18, pp. 2753–2756, 2006.

[35] D. L. Rabosky, "LASER: A maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies," *Evol. Bioinf.*, vol. 2, no. 4, pp. 247–250, 2006.

[36] J. Leskovec and C. Faloutsos, "Scalable modeling of real graphs using Kronecker multiplication," in *Proc. 24th Int. Conf. Mach. Learn.*, Jun. 2007, pp. 497–504.

[37] D. M. Chickering and D. Heckerman, "Efficient approximations for the marginal likelihood of incomplete data given a Bayesian network," in *Proc. 25th Int. Conf. Uncertainty Artif. Intell.*, Aug. 1996, pp. 158–168.

[38] I. Bezáková, A. Kalai, and R. Santhanam, "Graph model selection using maximum likelihood," in *Proc. 23rd Int. Conf. Mach. Learn.*, Jun. 2006, pp. 105–112.

[39] S. Jian, X. U. Ke, M. Song, and X. Zhan, "Credibility evaluation method of domestic IP address database," *J. Comput. Appl.*, vol. 34, pp. 4–6, May 2014.

[40] C. Ai, B. Chen, L. He, K. Lai, and X. Qiu, "The national geographic characteristics of online public opinion propagation in China based on WeChat network," *GeoInformatica*, vol. 22, pp. 311–334, Apr. 2018.

[41] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 10, Oct. 2008, Art. no. P10008.

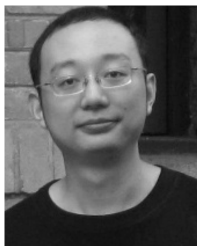
JIAN DONG is currently pursuing the master's degree with the National University of Defense Technology, China. His main research interest includes the modeling and simulation of the propagation of unconventional online public opinion.

BIN CHEN received the Ph.D. degree from the National University of Defense Technology, China, in 2010, where he is currently a Research Associate with the College of Information Systems and Management. He has hosted three projects of the National Natural Science Foundation of China in the field of emergency management of non-conventional emergencies. His research interests include emergency management of unconventional emergencies, modeling and simulation of network public opinion, and multi-agent modeling and simulation.

CHUAN AI received the master's degree from the National University of Defense Technology, China, in 2016, where he is currently pursuing the Ph.D. degree. His main research interest includes the modeling and simulation of the propagation of unconventional online public opinion.

PENGFEI ZHANG, photograph and biography not available at the time of publication.

XIAOGANG QIU, photograph and biography not available at the time of publication.



LINGNAN HE received the Ph.D. degree from Sun Yat-sen University, China, in 2012, where he is currently a Lecturer with the School of Communication and Design. His research interests include social psychology, communication, and public relations psychology, network public opinion analysis, new media behavior research, and user experience.