The Modeling of WeChat Propagation Network Based on Geographic Information

Chuan AI
College of Information System and Management, National University of Defense Technology Changsha China 410073
aichuan@nudt.edu.cn

Bin CHEN*
College of Information System and Management, National University of Defense Technology Changsha China 410073
nudtcb9372@gmail.com

Lingnan HE
School of Communication and Design, Sun yat-sen University Guangzhou China 510006
lingnan.he@qq.com

Kaisheng LAI
School of Communication and Design, Sun yat-sen University Guangzhou China 510006
kaishenglai@qq.com

Xiaogang QIU
College of Information System and Management, National University of Defense Technology Changsha China 410073
michael.qiu@139.com

Liang LIU
College of Information System and Management, National University of Defense Technology Changsha China 410073
nudt_ll@163.com

ABSTRACT

WeChat becomes the China's most widely used online social media application because of its convenience and simplicity. This paper is based on the records of HTML5 web pages that spread in WeChat. The propagation network G is constructed with these records. Then, according to the city where a Web user is, the network is divided into small networks within the city G_i (In-city Networks) and the networks G_{ij} between any two cities (Out-city Networks). Based on the analysis of G_i and G_{ij}, a WeChat propagation net-work model with geographic information is established. This net-work model has the characteristics of WeChat propagation and geographic information. This work can support the online opinion propagation researches with geographical characteristics.

CCS CONCEPTS

•**Networks → Network measurement;**

KEYWORDS

WeChat application, HTML5 webpage, propagation Network, Geographical characteristics

1 INTRODUCTION

With the development of online social networks, the influence of online public opinion is growing. The characteristics of privacy of the WeChat network have brought great difficulties to the supervision of public opinion.

With a large number of users, rich data and low-cost measurement, WeChat network analysis has advantages on traditional sociological measurement and analysis. With the classic social network analysis theory and technology applied to the WeChat network, we can get more accurate results on the social network. Through the network modeling the nature and characteristics of the network can be grasped and deeply understand. And then the support and basis for network supervision is provided.

From the perspective of space network, there are two most prominent regional scientific research based on the mobile communication data. The first one is that the gravity law could be used to fit the density of the interactions between the urban centers [1, 2]. Another one is about community detection, some studies [3, 4] found that the communities detected correspond with the administrative boundaries extremely well.

At the same time, a conclusion is found that there are more internal interactions in administrative units than that between two units in the research of this paper. Thus, for a vast country like China, the modeling of social media networks can be based on administrative units. That is, firstly a user interaction network within a single ad-ministrative unit (such as a city) is established, and then the interaction network between any two cities is considered. The network is not traditionally considered as a whole.

A large number of researches have shown that the real world net-works almost have small world effect [5, 6, 7, 8], and scientists also find that the nodesfi degree of real world networks are almost subject to the power-law distribution [9, 7, 10, 11, 8]. From this point of view, the actual characteristics of each city's internal Inter-active Network is analyzed for the modeling of each cityfis network. And then the Interactive Network between any two cities is analyzed, two methods are taken to model this part, and the results are analyzed and compared with each other to find the more suitable method.

The remainder of this paper is organized as follows. Section 2 introduces the related works in the field of OSN and modeling of network. Section 3 introduces the HTML5 dataset while section 4 proposes the modeling methods. Section 5 gives detailed analysis

*Corresponding author.

EM-GIS'17, Redondo Beach, CA, USA

DOI: 10.1145/3152465.3152469

of the Interactive Network modeling results. In the end, this paper is concluded in section 6.

2 RELATED WORKS

The network of social media application users is called online social network (OSN). The measurement and analysis of online social network refers to: Mining and sorting the original data of OSN; Mining and extracting the structural features and user behavior characteristics of OSN. It is usually done through the theoretical methods and techniques of complex network, social network and data mining. WeChat is a kind of online social application. In recent years, the academic research on OSN is becoming more and more extensive, including network topology, user behavior, user privacy and security, system architecture, community mining and information dissemination [12].

The researched on community detection based on mobile communication data mainly involves four aspects. In addition to the city-based network mentioned in the Introduction part, there are person-al-based [13, 14, 15] based on pixel grid [4], based on mobile base station division Into the Tyson polygons [16, 17] . The different types of networks are determined by different data sets. The nodes of these networks are geocoded into the city based on consumption data, based on square pixels or cell based on the caller's geographic location. The communities that detected in these network are often clearly aligned with administrative divisions. It is these results that guide us to construct the network from bottom to top based on giving administrative units.

Some researches show that the regular network has a big cluster coefficient and big small average shortest path, and the random network has a small cluster coefficient and a small average shortest path. In 1998, the WS network constructed by Watts and Strogats, and the NW network given by Newman and Watts [5, 18], had a large cluster coefficient and a small average shortest path. Later, the physicists put the two statistical features, large cluster coefficient and the small average shortest path together known as the small world effect. With this effect, a network is a small world network. At the same time, scientists also find that the nodes of a large number of real networks are subject to a power-law distribution, where the degree of a node is the number of adjacent nodes that the node has, or the number of edges associated with that node. The node degree is a power-law distribution. That is, the relationship between the number of nodes with a certain degree and the specific degree can be approximated by a power function. These findings inspire us to consider whether the interactive network within a single administrative unit also has a small world or scale-free effect.

3 DATASET AND TOPIC

HTML5 is a web-based programming language that supports multimedia on the mobile terminal.

The data set is derived from Fibonacci Consulting Co. LTD. HTML5 is spread among Moments, WeChat Subscrip and friends. Once a page is clicked, the server will generate and save a record. The record includes the ID of the clicker, the ID of the sharer, the time, the page ID, and the IP of the clicker. Each user has a unique ID, when he click to browse the page, his ID is recorded as the clickerfis ID, when he shared the page was click to browse, his ID in the record is record as the sharerfis. The historical data is collect-ed by Fibonacci as the data sets used in this article.

Table 1: Explanation of each part of the record

Name of parts	Explanation	Parts of record
SourceID	Sharer's	ID VhVMFZEr
ViewID	Clicker's ID	RwZZc4ke
PageID	Webpage ID	a70842c6-6847-404c -be56-6fb1b389228
IP	Clicker's IP	113.108.11.52
Click time	Click time	2016-01-14 00:00:00

The data set for this article is a 45-day history records for 2016.01.14-2016.02.27. Table 1 gives detailed explanations of five parts of this record.

Here is a record of data set:

$a70842c6-6847-404c-be56-6fb1b389228, VhVMFZEr, RwZZc4ke$, 113.108.11.52,, $2016-01-1400:00:00, switchpage$, "$ctp$" : 4, "$staytm$" : 2397, "$tgp$" : 5, "$ttp$" : 14

Through the IP library query we can get the corresponding geographical location. Song Jian [19] analyzed several major domestic and international IP address library, including foreign IP2Location Lite, GeoLite2 domestic $Pure$ IP Address Database, $TaobaoIP$ Address Database, $SinaIP$ Address Database, and Baidu IP Ad-dress Database. They compared coverage rate and accuracy of the four domestic IP address databases. They got the conclusion that the domestic $TaobaoIP$ Address Database had the highest credibility. Because the IP address library is difficult to test and verify, and the lack of information about building method of these two IP address databases, we cannot get credibility of $IP2Location$ Lite and $GeoLite2$. We use $TaobaoIP$ Address Database to address the geographical location of the IP address in our work.

At the same time, Song Jian et al.'s [19] work found that the lower the administrative level is, the lower the accuracy of IP address library is. In order to keep the accuracy, cities are taken as the basic units. All the administrative units under other provincial administrative units directly are regarded as basic units in our work. These units include sub-provincial cities, prefecture-level cities, county-level cities and autonomous counties. In our work, we also regard the four municipalities and two SARs as basic units in network, although they are provincial administrative units.

Thus there were 380 cities in the data. In particular, in this data set, due to the number of users in Shenzhen, the amount of intracity interaction is too large. It accounting for about half of the amount of interactions, which is inconsistent with the actual situation, so this factor will be removed to consider the WeChat interaction network. So, actually there are 379 cities in our data.

4 METHODS

This paper constructs a WeChat interactive network from the history records of HTML5 web pages. Through the $IPLibarary$ query, each user ID can be located. So all users could be divided into 379 groups which corresponded to 379 cities. Then the WeChat network is divided into the propagation networks within each city,

and the propagation networks between any two cities (we simply named them Infi?!city Networks and Out-city Networks). Through the analysis of Infi?!city Networks and Out-city Networks, the regular pattern of nodesfi degree distribution is found. Based on that, the Infi?!city Network model of each city, and the Out-city Network model are established.

4.1 Network Definition

Based on the information propagate relationship of each record in the data set, it is possible to construct a edge from one user to another, that is, if the a user has viewed the web page shared by the another, there is an undirected edge between them. All users and edges constitute an undirected network.

$$V = \{v_i | i = 1, 2.....N\} \tag{1}$$

$$E = \{e_j | j = 1, 2.....m\} \tag{2}$$

$$G =< V, E > \tag{3}$$

As there are usersfi IP the data in the data, so it is possible to locate the user's city, accordingly, the user node is divided into 379 sets, resulting in 379 In-city Network and C_{379}^2 within the interactive network and intercity interaction network.

City Interactive Network: V_i is the user node set of city i , and E_i is the edge set of the city.

$$V = \{V_i | i = 0, 1, 2.....378\} \tag{4}$$

$$G_i =< V_i, E_i >, i = 0, 1, 2.....378 \tag{5}$$

E_{ij}, G_{ij} is the edge set and Out-city Network of interaction of users between city i and j

$$E_{ij} = \{e_{ij} | e_{ij} = v_1 \rightarrow v_2, v_1 \in V_i, v_2 \in V_j\} \tag{6}$$

$$G_{ij} =< V_i, V_j, E_{ij} >, i, j = 0, 1, 2.....378, i < j \tag{7}$$

Let:

$$E^1 = \{E_i | i = 0, 1, 2.....378\} \tag{8}$$

$$E^2 = \{E_{ij} | i, j = 0, 1, 2.....378, i < j\} \tag{9}$$

$$G^1 = \{G_i | i = 0, 1, 2.....378\} \tag{10}$$

$$G^2 = \{G_{ij} | i, j = 0, 1, 2.....378, i < j\} \tag{11}$$

Thus $E =< E^1, E^2 >$ Th$G =< G^1, G^2 >$. G^1 is the set of in-city Networks, and G^2 is the set of Out-city Networks. E^1 is the set of all edges of In-city Net-works, and E^2 is the set of all edges of Out-city Networks.

The division of the network is completed. Figure 1 shows the city nodesfi size histogram of 379 cities.

It can be seen from the figure that the citiesfi nodes size varies greatly. The nodesfi size of largest city can reach more than 600 thou-sands, and that of small city is less than one thousand. So the nodesfi size of the city should be modeling is one of the important factors to consider.

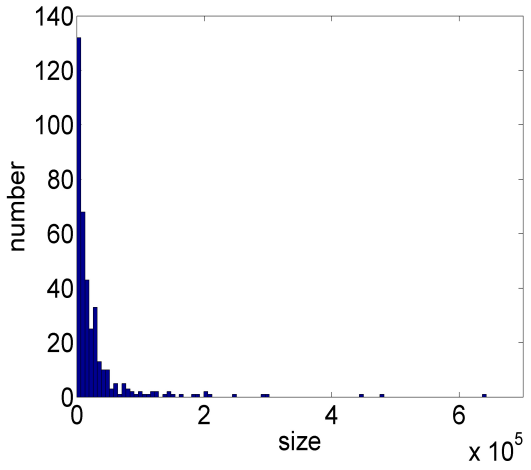

Figure 1: City scale histogram.

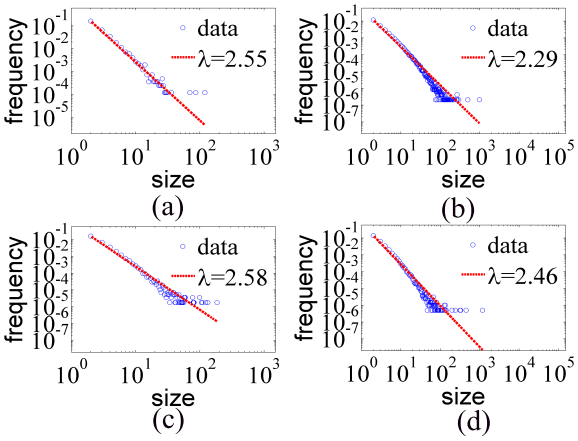

Figure 2: Nodes degree distribution of (a) Hani-Yi Autonomous Prefecture of Honghe. (b) Beijing. (c) Changchun. (d) Chengdu.

4.2 Network Analysis

The In-city Networks are analyzed first. Figure 2 (a) (b) (c) (d) are nodesfi degree distribution of four randomly selected cities. Through the analysis, it can be found that the nodesfi degree distribution of the interactive network in the city obeys the power-law distribution. The power-law coefficients λ ranges from 2 to 5. As the network is divided into cities, so the range of power-law coefficients obtained by fitting is relatively large.

Figure 3 shows the degree distribution for all citiesfi In-city Network, and their fitting lines. Their average power law coefficient is 2.761.

The vertical and horizontal coordinates are city numbers of 379 cities, two citiesfi coordinates corresponds to a gray point in the figure. Here we take the logarithm of m_{ij}, size of the set E_{ij}, and then normalized

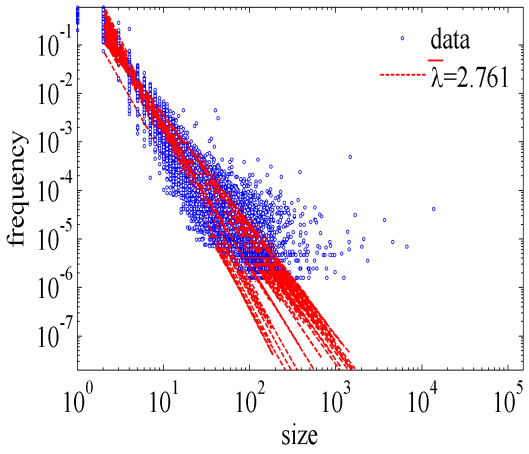

Figure 3: Degree distribution of all cities and their fitting line.
Notes: $\overline{\lambda}$ ***here is the average value of power-law coefficient of 379 cities.***

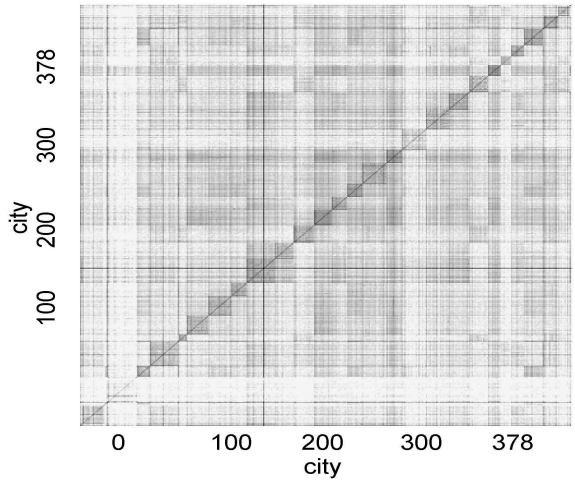

Figure 4: The density of the connection between the city and the city.

$$g_{ij} = \frac{\log(m_{ij})}{\max(\log(m_{ij}))} \tag{12}$$

Draw the gray value figure as shown in Figure 4. The darker points in the graph represent the bigger amount of interaction between the two corresponding cities.

It can be seen from analysis of figure 4 that color of right diagonal is obvious relatively deep, which means that the city's In-city inter-actions are significantly much more than that in Out-city Networks. This also means that the way to build an integrated interactive net-work based on the city is correct. At the same time, due to the city number, the city number of cities that belong to a provincial administrative unit is adjacent. It can be seen from the figure that the amount of interaction among cities belong to a province of nodes is relatively large, forming relatively deep color squares on the diagonal.

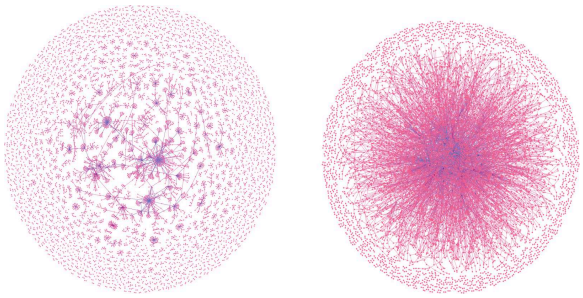

Figure 5: The original network and generated network of one city.

4.3 Network Modeling

The bottom-to-up method is used to model the WeChat propagation network. That is, the In-city Network is modeled and constructed based on the original network. And then modeling the Out-city Networks based on the interactions amount between any two cities.

4.3.1 In-city Networks. According to the analysis, it is found that the nodes degree of the In-city Network of each city is subject to the power-law distribution. Therefore, we can model the In-city Network of each city according to the nodesfi size N_i, $i = 0, 1, 2.....378$ and the power law coefficient λ_i, $i = 0, 1, 2.....378$ of the network.

Networkx package in python is used to randomly generate the node's degree sequence S_i based on the nodesfi size and power-law coefficients, and then generate the network according to the sequence

$$< N_i, \lambda_i > \xrightarrow{NetworkX\ \text{Tool}} S_i \tag{13}$$

$$S_i \xrightarrow{NetworkX\ \text{Tool}} G_i^{'} \tag{14}$$

The nodes set corresponding to network $G_i^{'}$ is: $V_i^{'}$, $i = 0, 1, 2.....378$.
And the corresponding edges set is: $E_i^{'}$, $i = 0, 1, 2.....378$.

Gephi is used to draw the original network and the generated net-work of Hani-Yi Autonomous Prefecture of Honghe. We can com-pare the original network with the generated network as shown in figure 5.

Take this city as an example, it can be found that $G_i^{'}$ show some characteristics of G_i, however, there are some characteristics that $G_i^{'}$ cannot present. With analysis, it can be found that the city's In-city Network is a typical star radial network. That is, there is core area exists in the network, and the nodes in the core area have more edges, and the generated network also has such characteristics. But there are some differences between them. The main difference is that the core of the generated network forms a community as a whole, and the star-shaped radial pattern is obvious. However, there are several smaller cores in the original network, there is not a core area as a whole. The original network formed

a number of core areas with different size, which may be the characteristics of WeChat propagation network.

4.3.2 Out-city Networks. The propagation network should be modeled with the characteristics of original network without being limited to the original net-work. In addition to build In-city Network, the Out-city Network of interactions between any two cities also need to be modeled. The nodes size is confirmed when modeling the In-city Networks, and the edge number of the network between any two cities can also be calculated. Thus when building the network, we must ensure the number of edges, and each edge links two nodes in corresponding two cities. The problem is how to choose the two nodes of each edge in the two city. We propose two methods to choose nodes.

The first method is similar to the ER network construction method: for each edge that need to add, one node is chosen by a same percentage in each city to generate the edge.

The percentage of choosing $Node_i$for one edge is:

$$\prod_i = \frac{1}{N_j}, \quad i = 0, 1, 2.....N_j, j = 0, 1, 2.....378 \tag{15}$$

N_j is the nodes size of city j . The second method is a preferential attachment mechanism similar to a scale-free network. The preferential attachment here refers to the fact that the newly added nodes are more likely to be connected to the hub nodes with higher degree. This phenomenon is called "Rich get Richer" or "Matthew effect" as new articles are more likely to refer to those widely cited important literature. Based on the characteristics of preferential attachment, Barabasi and Albert proposed the BA scale-free network model.

The second method for each edge to be added: in proportion to the probability of nodesfi degree in the network of each city, one node is selected to generate an edge.

The percentage of choosing $Node_i$for one edge is:

$$\prod_i = \frac{k_i}{\sum k_i} \tag{16}$$

$k_i$is the degree of $Node_i$. This makes the overall net-work shows more scale-free characteristics compared to the first method. $E^{'}_{ij}$, $G^{'}_{ij}$ are the set of interactive edges and net-work between city $i$and$j$:

$$E^{'}_{ij} = \{e^{'}_{ij} | e^{'}_{ij} = v_1 \rightarrow v_2, v_1 \in V^{'}_i, v_2 \in V^{'}_j\} \tag{17}$$

$$G^{'}_{ij} =< V^{'}_i, V^{'}_j, E^{'}_{ij} >, i, j = 0, 1, 2.....378, i < j \tag{18}$$

Can determine the size of the city's nodes in the construction of a single city network has been identified, and the city between the two sides of the user interaction can be obtained with the number of statistics, so in modeling, the first need to ensure that the two cities Network edge is determined, each side is linked to two city users. Therefore, this paper uses two methods to model this part of the network.

4.3.3 WeChat propagation network. With the In-city network and the Out-city network, we can integrate and generate the whole WeChat propagation network.

Let:

$$E^{1'} = \{E^{'}_i | i = 0, 1, 2.....378\} \tag{19}$$

$$E^{2'} = \{E^{'}_{ij} | i, j = 0, 1, 2.....378, i < j\} \tag{20}$$

Table 2: The comparison of original and generated networkfis parameters

Networks	Size	Average Degree	Coefficient
Original	38527609	1.463	3.15
Generated 1	38527609	1.3	2.74
Generated 2	38527609	1.45	3.08

$$G^{1'} = \{G^{'}_i | i = 0, 1, 2.....378\} \tag{21}$$

$$G^{2'} = \{G^{'}_{ij} | i, j = 0, 1, 2.....378, i < j\} \tag{22}$$

So, we have $E^{'} =< E^{1'}, E^{2'} >$ Th$G^{'} =< G^{1'}, G^{2'} >$. $G^{1'}$ is the set of In-city Net-works, and $G^{2'}$ is the set of Out-city Networks, $E^{1'}$ is the edge set of In-city Networks, $E^{2'}$ is the edge set of Out-city Networks. Finally, we get the WeChat Propagation network$G^{'}$. For the convenience of description, in the following description of this paper, the method of generating the Out-city Network by the first method is called method one. The method of generating the Out-city Network by second method and finally generating the WeChat network is called method two.

5 RESULTS ANALYSIS

Firstly, the effect of the two methods to generate Out-city Network on the overall network is analyzed. A simple analysis of the original network and the generated network is shown in table 2. The two generated network and the original network are the same on the scale. The original network is the largest on the average degree, the method is second, the network is small. The power law coefficient is also the method two is close to the original network, but the method two generation network power law coefficient is small.

As can be seen from table 2, the two methods to generate the Out-city network have a certain impact on the resulting network. Com-pared with the original network, the network generated with the second method has closer parameters with original network.

The nodes size distributions of three networks are analyzed as shown in Figure 6, 7, 8. It can be seen that generated network with second method is significantly closer with the original network in the node's degree distribution. And the percentage of nodes with degree on ten to one hundred is obviously large in generated network with first method, showing a raised spike.

It can be guessed that there are obvious problems in the network generation process. It can be guessed that because, the rules of edge addition is not correct in process of the Out-city Network construct-ing, resulting in a large difference between the generated network and the original network. In order to further confirm the conjecture, all the Out-city Networks are fused, and the nodes degree distribution is shown in figure 9.

It can be seen that there is a great difference with the scale-free network as shown in figure 9. In the interval of less degree, the distribution of the degree can be characterized by the power law distribution, and the increment is shown on one hundred to one thousand, obviously, this is a kind of characteristic that does not conform to the characteristics of scale-free. It can also be seen from the nodesfi degree distribution of network finally obtained

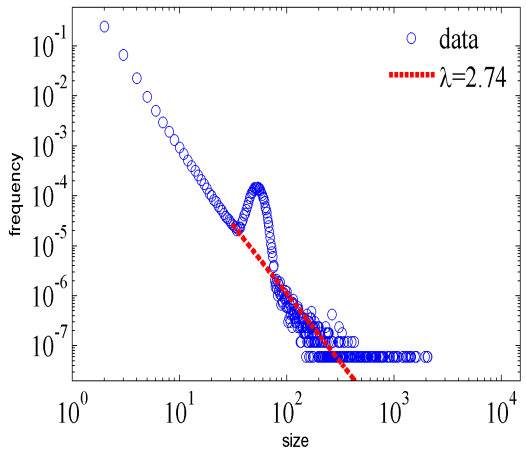

Figure 6: The original network and generated network of one city.

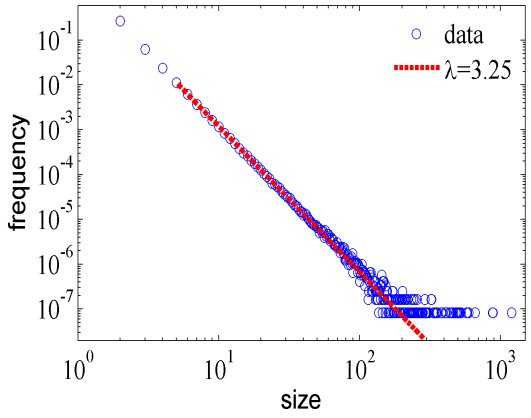

Figure 7: Nodesfi degree distribution of generated network with method two.

as shown in figure 6, there is significantly difference between the generated network and the original network.

The network generated with second methods is relatively closer to the original network, suggests that the model is more reasonable.

However, there are obvious problems in the whole modeling pro-cess.

There is still a certain difference between the network generated and the original network during the In-city network modeling process for In-city Network. The original network of a single city is more like a big group linked with a number of small group formed by relatively denser links. The generated network is a typical kind of star-radial network, which forms a network with obvious core. This is also a very important issue to be solved in the future.

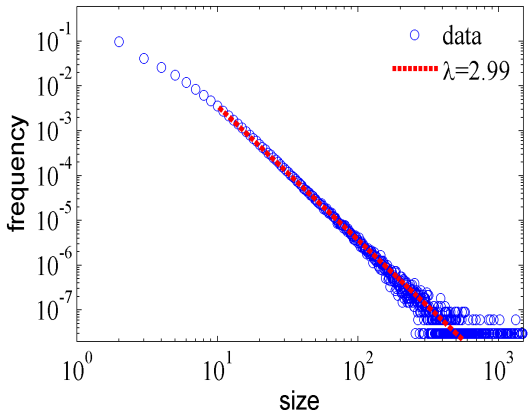

Figure 8: Nodesfi degree distribution of original network.

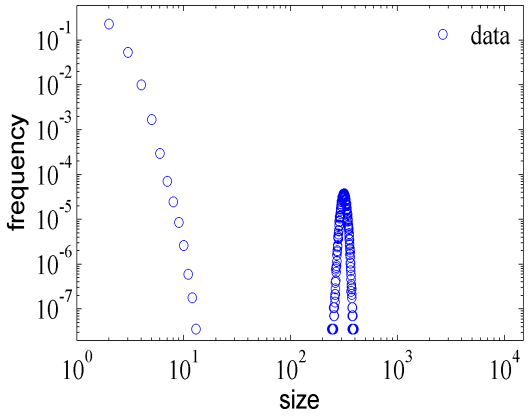

Figure 9: Nodesfi degree distribution of generated Out-city network with second method.

6 CONCLUSIONS

By constructing and dividing WeChat propagation network, the In-city Networks and Out-city Networks are built and then the whole WeChat propagation network is modeled and constructed. We pro-pose two methods to generate the edges of Out-city Network, and find that there are obviously shortages in first method, the network generated with second methods is consistent with original network basically.

The modeling method of WeChat propagation network can provide the mechanism of stochastic generation of WeChat propagation networks, which is very important for the research for public opinion communication. In particular, the research on propagation of public opinion in the past is mostly conducted with a step-forward or sequential progression method, and the work of this paper can provide the research of public opinion propagation with the net-work model containing spatial information.

Based on the work of this paper, the simulation of public opinion propagation, each node has a specific geographical location

information to the prefecture-level, which has an important role in public opinion propagation research focused on the geographical information.

However, there are some areas in the work of this paper worth improving. First of all, the city's In-city Network modeling parameters considerated are relatively simple, only the nodesfi scale and degree distribution power-law coefficients, without consideration of clustering coefficients, network internal associations and other factors. This is also reflected in the differences between the two net-works in figure 5. Secondly, the modeling of Out-city network needs more thorough research work, at present, simply considering the scale-free characteristics. There should be more detailed re-search and a more detailed model should be constructed.

In general, the current modeling method can basically generate the WeChat propagation network based on geographic information. This plays an important role in the research of WeChat social media network, it is also the base of the research on public opinion communication on this kind of social network. It is of great significance for the research of public opinion communication with geographical information.

7 ACKNOWLEDGEMENTS

This study is surpported by National Key Research Development (RD) Plan under Grant No. 2017YFC080300 and the National Natural Science Foundation of China under Grant Nos. 71673292, 61503402 and Guangdong Key Laboratory for Big Data Analysis and Simulation of Public Opinion. We also thanks Fibonacci Consulting Co., Ltd for the data set.

REFERENCES

[1] K., Gautier, C., Francesco, R., Carlo, and DB., Vincent. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, 07(L07003), 2009.

[2] C, Roth, SM, Kang, M, Batty, and M, Barthélemy. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS One*, 6(1):e15923, 2011.

[3] V, Blondel, G, Krings, and I, Thomas. Regions and borders of mobile telephony in belgium and in the brussels metropolitan zone. *Brussels Studies*, 42, 2010.

[4] C, Ratti, S, Sobolevsky, F, Calabrese, C, Andris, J, Reades, M, Martino, R, Claxton, and SH, Strogatz. Redrawing the map of great britain from a network of human interactions. *PLoS One*, 5(12), 2010.

[5] WATTS, D. J. and STROGATZ, S. H. Collective dynamics of 'small-world' networks. *Nature*, p. 440, 1998.

[6] F, LILJEROS, R, EDLING C, and A, AMARAL L. The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001.

[7] H, EBEL, I, MIELSCH L, and BORNHOLDT S, . . Scale-free topology of email networks. *Physical Review E Statistical Nonlinear Soft Matter Physics*, 66(3), 2002.

[8] S, Milgram. The small world problem. *Psychology Today*, 2(1):185–195., 1967.

[9] M, FALOUTSOS, P, FALOUTSOS, and C, FALOUTSOS. On power-law relationships of the internet topology. *Acm Sigcomm Computer Communication Review*, 29(4):251–262, 1999.

[10] L, BARABASI A and ., ALBERT R. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[11] R, ALBERT, H, JEONG, and BARABASI A L . ., ():. Diameter of the world wide web. *Nature*, 401(6749):130–131, 1999.

[12] KE, X.U., ZHANG, S., HAO, C., and HAI-TAO. Measurement and analysis of online social networks. *Chinese Journal of Computers*, pp. 29–42, 2014.

[13] J, ONNELA, J, SARAMÄKI, J, HYVÖNEN, G, SZABÓ, M, DE MENEZES, K, KASKI, A, BARABÁSI, and J, KERTÉSZ. Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics*, 9(6), 2007.

[14] G, PALLA and T, VICSEK. Quantifying social group evo-lution. *Nature*, 446(7136):664, 2007.

[15] R, LAMBIOTTE, V, BLONDEL, C, DE KERCHOVE, E, HUENS, C, PRIEUR, Z, SMOREDA, and P, VAN DOOREN. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21):5317–5325, 2012.

[16] G, SONG, L, YU, W, YAOLI, and ., XIUJUN M. Discovering spatial interaction communities from mobile phone data. *Transactions in GIS*, 17(3):463–481, 2013.

[17] C, GUANGHUA, T, JEAN-CLAUDE, T, DAOQIN, S, LI, and L, YU. Uncovering regional characteristics from mobile phone data: A network science approach. *Papers in Regional Science*, 95(3):613–631, 2016.

[18] J, Newman M E and J, Watts D. Scaling and percolation in the small-world network model. *Physical Review E Statistical Physics Plasmas Fluids Related Interdisciplinary Topics*, 60(6 Pt B), 1999.

[19] JIAN, SONG, XU, MEINA, SONG, and XIAOSU, ZHAN. A method of assessing the credibility of domestic ip address libraries. *Computer applications*, z2(15):4–6, 2014.