

Correlation and Stability Study of Centrality in Social Network

Tong Jingwen^{1,2}

1. Chongqing Collaborative Innovation Center for Information Communication Technology
2. Electronic Information and Networking Research Institute, Chongqing University of Posts and Telecommunications
Chongqing, China
449697730@qq.com

Luo Jiangtao^{1,2}

1. Chongqing Collaborative Innovation Center for Information Communication Technology
2. Electronic Information and Networking Research Institute, Chongqing University of Posts and Telecommunications
Chongqing, China
Luojt@cqupt.edu.cn

Abstract—Social Network Service (SNS) includes QQ, Facebook, Twitter, Blog, etc. Studies show that in the complex network, 20% of nodes with degree accounted for 80% of total degree. So the importance of each node is very different in the network. Many researchers dedicated to studying of complex network centrality, such as degree, betweenness and closeness centrality, and finding out the most important nodes (topN) in the network to implement the regulation and maintenance of the whole network. However, the actual network is changing constantly and the network topology is not machine-parsable. In this paper, using the BA model to verify the stability of the network centrality in the process of changing network topology. It showed that the degree, betweenness and closeness centrality remain stable in the process of network changing, and according to the topN of three kinds of centrality concluded the relationship between them.

Keywords-social network; network centrality; BA model; stability; correlation

I. INTRODUCTION

Social network is a part of the complex network, so it has the basic characteristics of a complex network, such as the node degree and degree distribution, average path length and clustering coefficient^[1]. They are basic studies of the small world properties and scale-free networks characteristics. Network model (WS model) which has the characteristics of a small world is to add a small amount of randomness in the regular network. Although WS can reflect characteristics of the small world of the real network, it's very rare nodes have a large number of connections, on the contract, a majority of nodes only have a few connections. Aiming at this phenomenon, scale-free network model (BA) was proposed by Barabasi and Albert which demonstrates the phenomenon. BA model has two important features: growth and preferential attachment^[2].

The importance of each node in the network is very different. Zhang Ke^[3] through the complex network theory and analytical methods studies the BBS reply networks, which show that the closeness centrality has a great influence on the paces of information carrying. Gao Chao^[4] analyzes the ability of different immune strategies on the inhibition of the virus in the interactive email model, especially for the immune strategies based on information of betweenness, it shows that the node betweenness

immune ability to curb the spread of the virus at minimal cost. Wu Xuguang^[5] explores the relationships of multiple centralities with factor analysis, experiment results show that the widely used five centralities in scale-free complex networks are divided to two classes: 1) degree, betweenness and eigenvector that mainly reflect the node influences with others; 2) closeness and eccentricity related to network topology. Wang Xiaoguang^[6] uses betweenness as indicators based on the comprehensive analysis of degree influence factor, proposed an evaluation method that betweenness affect the spread, it also giving the betweenness calculation method based on contribution. Guan Jianfeng^[7] compares the betweenness, degree, closeness and eigenvector centrality, and draw a conclusion that betweenness centrality than other centralities has a better cache performance.

In the real world, networks are continuously changing, however, when the network topology starts to change, the network centrality unchanged is also a point which is worthy to study. Inspired by the BA model, this paper used the BA model algorithm to verify the stability of the network centrality. It's shown that with the expanding of network size, the network centrality remains stable by recreating the BA model algorithm, and get the relationship between the three kinds of centralities.

II. BA SCALE-FREE NETWORK MODEL

A. Two important Characteristics of the BA Model

BA model has two important features:

1) *Growth*: Network scale is expanding. Such as the Internet has a large amount of data generated every day and has new articles of science and technology every month.

2) *Preferential attachment*: New nodes are more inclined to connect large node degree. This phenomenon is called "the Matthew effect" or "the Rich get richer". For example, Taobao consumers tend to choose the evaluation of high commodity in similar products, new articles being more inclined to cite the important documents of high degree.

B. The Algorithm of BA Model

Based on the characteristics of the BA model, the formation algorithm of the BA model is as follows.

1) *Growth*: Begin with a network containing m_0 nodes. Add one new node into the network every time, and connect it with m old nodes, where m is less than or equal to m_0 .

2) *Preferential attachment*: The probability Π_i of connecting the new node to an old one i is given by

$$\Pi_i = \frac{k_i}{\sum_i k_j}. \quad (1)$$

Where k_i is the degree of i , and k_j is the degree of another old node j ^[8].

In fact, the structure algorithm of the BA model is evolved from the Price model algorithm, and it is the exception of the Price model. Specific implementation is as follows:

① given an initial strongly connected graph with m_0 nodes, and added the node ID of each edge points to the Array.

② given the parameter $p \in [0, 1)$. For $t=1, 2, \dots, N-m_0$ to execute the follow steps:

- Generate a random number completely $r \in [0, 1)$;
- If $r < p$, then completely randomly choose one of the elements in the Array;
- If $r \geq p$, then completely randomly choose a node;
- After the previous three steps m times, then add new nodes that point to the selected m nodes and m edges, and add the node ID of the m nodes to Array.

Where $m=a$, $p=0.5$, and we can get the BA model by letting all the directed edges turn to undirected in the network topology of the parameters generated^[1].

C. The pseudo-code of the BA Model Algorithm

TABLE 1. is the pseudo-code of BA model algorithm.

TABLE I. BA MODEL ALGORITHM

Algorithm: the BA model
Input: the file that is network topology that includes nodeid and edges.
Precondition: file.open the file
Main program:
1: std::vector<string> nodeid, nodeid.push_back(line)
2: std::vector<string> array, array.push_back(to)
3: std::vector<string> nodes
4: $p=0.5$, for $t=1, 2, \dots, N-m_0$, then:
5: $r1=((\text{double})\text{rand}()/\text{RAND_MAX})*(b-a)+a$, where $a=0, b=1, r \in [0, 1)$
6: if $r1 < p$
$x=\text{nodeid}[\text{rand}()\% \text{totnode}]$, nodes.push_back(x)
7: else
$y=\text{nodeid}[\text{rand}()\% \text{totnode}]$, nodes.push_back(y)
8: after implement 5-7 m times (m is the number of edges of a node add to the original figure, where $m=2$)
array.push_back(x) or array.push_back(y).

III. THE DEFINITION AND CALCULATION OF NETWORK CENTRALITY

The undirected graph G is defined by a pair of sets $G: (V, E)$, where V is a non-empty countable set of nodes, and E is the set of edges; and from these the following concepts and definitions are obtained.

1) *Degree centrality*: The degree centrality $C_D(v)$ of node v is

$$C_D(v) = \frac{\text{deg}(v)}{n-1}. \quad (2)$$

Where $\text{deg}(v)$ is the degree of v , which is defined as the number of edges in the graph incident upon a node^[9].

2) *Betweenness centrality*: The betweenness centrality $C_B(v)$ of node v is

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\delta_{st}(v)}{\delta_{st}}. \quad (3)$$

Where δ_{st} is the number of shortest paths from s to t , and $\delta_{st}(v)$ is the number of shortest paths from s to t that pass through a node v ^[10].

3) *Closeness centrality*: The closeness centrality $C_C(v)$ of node v is defined as the mean shortest path between v and all other nodes reachable from it

$$C_C(v) = \frac{1}{\sum_{s \in V \setminus v} d(v, s)}. \quad (4)$$

Where $d(v, s)$ is the shortest path between node v and node s ^[9].

IV. EXPERIMENTAL DESIGN AND IMPLEMENTATION

A. Design scheme

Fig. 1 is a BA model with 1000 nodes and 1997 edges generated by BRITE topology generator^[11]. These parameters are set to: $m=2$, Max BW=1024, Min BW=10. First, calculate the betweenness, degree and closeness centrality of each node in the network topology, choosing the top 15 nodes of three centralities. In the Fig.1, the different circles sizes represent different nodes' centrality in the network, the bigger circle is the bigger centrality. Then, adding new nodes to the topology via the structure algorithm of the BA model, each time adds 1000 nodes and 2000 edges. Calculate the top15 of three centralities in the network after adding new nodes and edges. Finally, it will turn to be a BA model topology which has 10000 nodes and 19997 edges.

According to the calculated the top15 of three centralities, getting the relationship between the kinds of three centralities via using Matlab simulation.



Figure 1. The BA Model Topology of 1000 nodes.

B. Conclusion and analysis

Experimental results are showed in Fig .2 and Fig .3. Fig .2 shows the following three conclusions: (1) In the progress of the network nodes increased from 1000 to 10000 nodes, the betweenness, degree and closeness centrality of the top 10 nodes basically remains the same, just in a different rank order. (2) With the increase of network scale, there will always be new node ID appears among the top15, which is normal. Such as in social networks, with the increase of time and the expansion of network scale, one of nodes will have bigger and bigger influence to other nodes, so it will become more and more important in the network. (3) In general, in the top10 nodes, the degree and betweenness centrality always between 0 to 40, while the closeness centrality between 0 to 20. It shows the closeness centrality of network, comparing to the degree and betweenness centrality has better stability.

Fig .3.1 shows the connections between betweenness and degree. With the increase of node degree, betweenness also shows a trend of growth. The nodes with smaller betweenness and degree are located in the edge of the network; while the bigger ones are located in the center of the network. The nodes which have smaller degree and higher betweenness play a role of bridge. Fig .3.2 shows the relations between closeness and degree. The nodes with higher degree also have higher closeness, which means the nodes more close to the network center, have less chance to be controlled by others. So nodes with lower degree have lower closeness, which depends on others forwarding to get information.

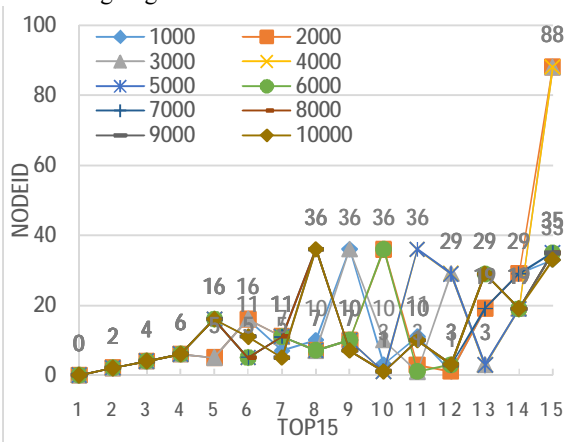


Figure 2.1. Betweenness Centrality Stability.

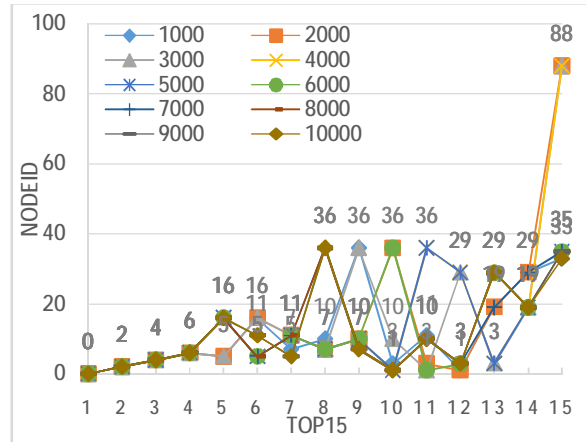


Figure 2.2. Degree Centrality Stability.

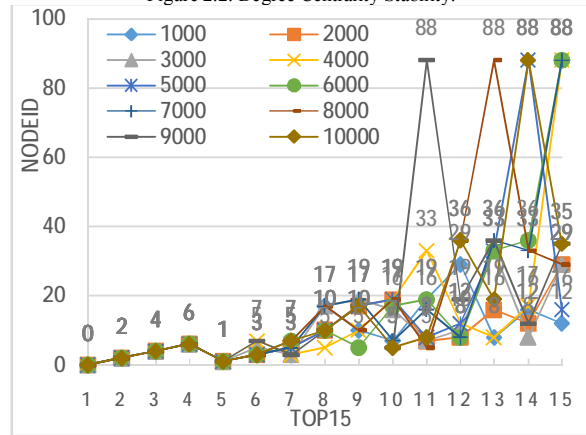


Figure 2.2. Closeness Centrality Stability.

Figure 2. The Stability of three Centralities.

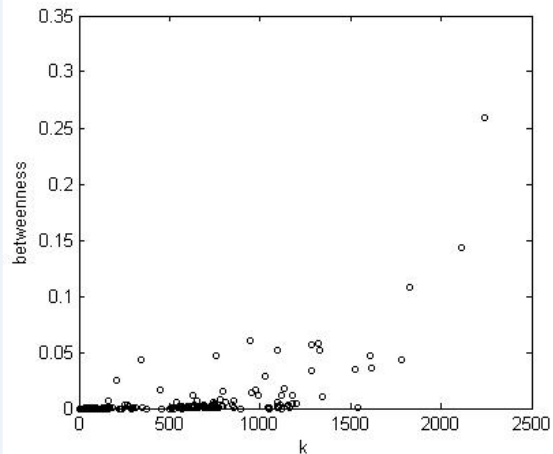


Figure 3.1. The Correlation between Degree and Betweenness.

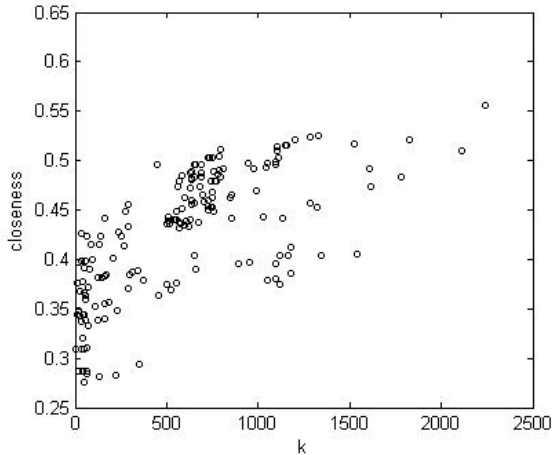


Figure 3.1. The Correlation between Degree and Closeness.

Figure 3. The Correlation of three Centralities.

V. CONCLUSION

When the network uses the BA model to extend, the degree, betweenness and closeness centrality in the network are stable. This paper also gets the correlation among three centralities. Therefore, in the real network, no matter how the web changes, through the most important nodes strengthen the management, maintenance of the network, and research on the security of the network can be done. For example, in the QQ group network through the most important nodes can strengthen the management of the QQ group. In the microblog network, the important information is sent from the important nodes in order to spread the information to the whole network quickly. By strengthening the security of the most important nodes to protect the whole network security in the mail network.

ACKNOWLEDGMENT

This work was supported in part by the Chongqing Municipal Application and Development Planning Project under GRANT cstc2013yykfA40006 and Program for Innovation Team Building at Institutions of Higher Education in Chongqing under GRANT KJTD201312.

REFERENCES

- [1] Wang Xiaofan, Li Xiang, Chen Guanrong, "Network science: an introduction[M]," Beijing: Higher education press, pp. 35-324, 2012.
- [2] Barabási A L, Albert R, Jeong H, "Mean-field theory for scale-free random network[J]," *Physica A: Statistical Mechanics and its Application*, 1992, 272(1): 173-187.
- [3] Zhang Ke, Li Hui, Qin Lijuan, et al, "Closeness centrality on BBS reply network[J]," *International Conference of Information Technology, Computer Engineering and Management Sciences*, 2011.
- [4] Gao Chao, Liu Jiming, Zhong Ning, et al, "Research on betweenness-based immunization strategy in email networks," *Computer Engineering*, Vol. 36, No. 5, March 2010.
- [5] Wu Xuguang, Zhang Mingqing, Han Yiliang, "Research on centrality of node importance in scale-free complex networks[J]," *Proceedings of the 32st Chinese Control Conference*, July 25-27, 2012, Hefei, China.
- [6] Wang Xiaoguang, Wang Feng, Li Shen, "Node importance evaluation of communication network based on betweenness influence matrix[J]," *Journal of Air Force Engineering University(Natural Science Edition)*, vol. 13, No. 5, 2012.
- [7] Guan Jianfeng, Quan Wei, Xu Changqiao, "The location selection for CCN router based on the network centrality[J]," *Proceedings of IEEE CCIS2012*.
- [8] Guo Chonghui, Zhang Liang, "An improved BA model based on the PageRank algorithm[J]," *IEEE*, 2008: 978-1-4244-2108-4/08.
- [9] Freeman, L. C, "Centrality in social networks: conceptual clarification[J]," *Social Networks*, pp: 215-239, 1979.
- [10] Freeman, L. C, "A set of measures of centrality based on betweenness[J]," *Sociometry* 40, 35-41, 1977.
- [11] Brite[EB/OL]. NSF grants CAREER ANI-0096045 and ANI-9986397, Alberto Medina, Anukool Lakhina, Ibrahim Matta, John Byers. [2015-05-28]<http://www.cs.bu.edu/brite/>