LETTER

An Approximate Flow Betweenness Centrality Measure for Complex Network

Jia-Rui LIU[†], **Shi-Ze GUO**[††], *Nonmembers*, **Zhe-Ming LU**[†a)], *Member*, **Fa-Xin YU**[†], *Nonmember*, *and* **Hui LI**[†], *Member*

SUMMARY In complex network analysis, there are various measures to characterize the centrality of each node within a graph, which determines the relative importance of each node. The more centrality a node has in a network, the more significance it has in the spread of infection. As one of the important extensions to shortest-path based betweenness centrality, the flow betweenness centrality is defined as the degree to which each node contributes to the sum of maximum flows between all pairs of nodes. One of the drawbacks of the flow betweenness centrality is that its time complexity is somewhat high. This Letter proposes an approximate method to calculate the flow betweenness centrality and provides experimental results as evidence.
key words: *complex network, centrality, flow betweenness centrality, approximate flow betweenness centrality*

1. Introduction

In the world, there are various systems. Every system is composed of elements, which are interrelated and interacted to form a whole. If we adopt a node to represent each element and utilize an edge to represent the contact between each pair of elements, a system can be viewed as a network. For example, the whole human society is a complex network with multi-level structures.

Complex network theory comes from the graph theory, and two famous Hungary mathematicians Erdös and Rényi have made great contribution to its development. Since they put forward a theory that is known as ER random graph model, they have controlled the study of complex network over 40 years. However, it has been found that many results that are obtained by calculating some actual network data depart from the random graph theory, so new network models are required to describe these real network characteristics. In the late 1990's, two pioneering works broke the framework of random graph theory. One is the article published in "Nature" where Watts and Strogatz proposed a small-world network model known as the WS model [1]. This model simulates many real-world networks that have high clustering coefficients and short average path lengths. Newman and Watts improved the original model and proposed the NW model [2]. Another is the article published in "Science" where Barabási and Albert pointed out that the degree distribution form of many real-world complex networks complies with the power law [3]. As there is no obvious characteristic length for power-law distribution, such networks are called scale-free networks. On this basis, Barabási and Albert established the BA model based on the growth and preferential attachment mechanism, and gave the numerical and analytical solutions [3]. These researchers' works have been recognized by academia, and a large number of scholars have joined the ranks of those working in complex networks. Lots of research works have been made in the empirical research, evolutionary models and network dynamics, which are related to physics, biology, social science, technological networks, engineering, economic management and many other fields.

In general, a complex network contains many nodes, but not all nodes share the same importance, some nodes may make more contribution to the network structure. If some important nodes in a network are removed, the network may suffer great damage or even cannot work at all. For instance, some diseases are spread through close contact between people. If we manage to find out what kind of people more likely become the medium of spreading disease and cure them, it will be easier to prevent the spread of disease. Therefore, for a harmful network, it becomes an important issue to find out the nodes with more importance and destroy them. On the contrary, for a helpful network, we can protect them to make the network safer.

The centrality measure is first proposed for social network analysis, and later widely used to other fields. Nowadays, it becomes a fundamental issue to find out the important nodes when analyzing a complex network. There have been many measures to evaluate the importance of a node in complex networks [4]. These measures can be classified into two categories. The first category thinks that the importance of a node is equivalent to the outstanding feature arisen when the node has connection with other nodes. In other words, this category adequately reflects the nodes' positional characteristic in the network, and defines the importance of a node by zooming the nodes' significance. The second category is based on a hypothesis that the importance of a node is equivalent to its destructivity. That is, if a node is removed from a network such that great damage will occur, then this node should be an important node in the network. The simplest centrality measure is degree centrality [5], i.e., a node with high degree is considered as

Manuscript received July 18, 2012.
Manuscript revised October 30, 2012.
[†]The authors are with the School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China.
[††]The author is with North Electronic Systems Engineering Corporation, Beijing 100083, China.
a) E-mail: zheminglu@zju.edu.cn (Corresponding author)
DOI: 10.1587/transinf.E96.D.727

an important node. Considering the traffic characteristics of a network, one can use the closeness centrality [6] that denotes the difficulty that a node has connection with other nodes. For a node with a smaller degree but playing a key role in the network connectivity, the betweenness centrality [7] can be adopted. The flow betweenness [8] ascertains the network geometry center by the way of flowing. The eigenvector centrality [9] is used to evaluate the impact on a node from other connected nodes. According to the participation level that a node takes part in different sub-graphs, the subgraph centrality [10] has also been proposed. If the network information is incomplete, then the random-walk betweenness centrality [11] can be utilized.

As we know, it is computationally expensive to exactly calculate betweenness, currently the fastest algorithm proposed by Brandes [12] has the time complexity of $O(NM)$ for unweighted graphs, where N is the number of nodes and M is the number of edges. In fact, in order to detect centroids in a network, it is unnecessary to compute the accurate betweenness centrality value of all nodes but only their relative betweenness value. Therefore, many scholars have been engaged in approximation approaches. Bader et al. [13] presented an adaptive sampling technique that significantly reduces the number of single-source shortest path computations for nodes with high centrality. Geisberger et al. [14] proposed a framework for unbiased approximation of betweenness. Gkorou et al. [15] proposed an approximation betweenness centrality for large and dynamically growing networks. Lee et al. [16] presented a method that efficiently reduces the search space by finding a candidate set of nodes and computes their betweenness using candidate nodes only.

Above-mentioned improvements are all designed for calculating betweenness centrality, which should perform time consuming computation of shortest paths. This Letter focuses on the flow betweenness centrality [8], which is redefined from the measure of betweenness centrality without considering whether the path between a pair of nodes is the shortest one or not. This measure calculates all possible paths connecting two nodes, and it is also known as the generalized betweenness centrality. If a network is with N nodes, then the flow betweenness of Node i can be defined as:

$$C_B(i) = \sum_{j<k} \frac{g_{jk}(i)}{g_{jk}} \tag{1}$$

where $C_B(i)$ is the flow betweenness of Node i, g_{jk} is the number of all possible paths between Node j and Node k, and $g_{jk}(i)$ is the number of all possible paths between Node j and Node k that pass through Node i. Although the flow betweenness measure can obtain more accurate network geometry center, it takes more time to deal with data especially when a network has a huge number of nodes. In this Letter, we first propose an approximate calculation method of the flow betweenness centrality, and then compare our measure with the flow betweenness centrality by analyzing a simple network. Furthermore, we show the effectiveness of our method by analyzing an example network.

2. Approximate Calculation Method

In fact, the flow betweenness of a node is a probability that the information created by other nodes flows through this node after a long time of free flowing. In other words, if we suppose a node in the network creates a message segment, then after some steps the message will reach any other node. In the flowing process, as long as the network is a connected graph, the message will definitely flow through a given node i. Here, we assume that the given node just receives information without any output, and thus it is easy to know the amount of information it has collected from other nodes. In other words, the flow betweenness of the given node i can be defined as the ratio of its collected information to the sum of information in the whole network. Therefore, our proposed method can be simply described as follows:
Step 1: Initialization. Define the amount of information for each node is 1;
Step 2: Information flowing. Suppose the degree of a node is k, then the information that flows to its each adjacent node is $1/k$. In this way, the whole information capacity is a constant value that equal to the number of nodes.

If the diameter of a network is D, we can think that after D steps, the information created by a node will be spread throughout the entire network.

Based on the above idea, we can get an approximate calculation result for each node's flow betweenness. If U_i stands for a set of k nodes connected with node i and λ_j equals $1/k$, then the above approximate calculation can be defined as:

$$C_i^{FM} = \sum_{k=1}^{D} \sum_{j \in U_i} \lambda_j C_j(k) \tag{2}$$

$$C_j(k) = \sum_{m \in U_j} \lambda_m C_m(k-1), m \neq i \tag{3}$$

In general, if a node is close to the network center, its corresponding centrality value will be big. Below we compare our approximate calculation results with the original measure by analyzing a simple network, which is generated randomly. The results are shown in Fig. 1 and Fig. 2 respectively.

From above two figures, we can see that the approxi-

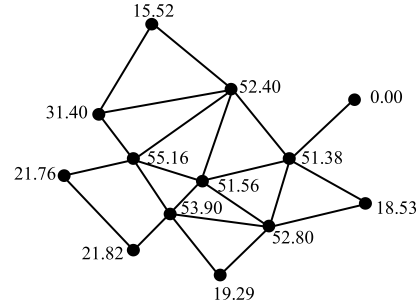

Fig. 1 Results of the flow betweenness centrality.

mate calculation also can accurately find the geometric center of the network. However, there is a little difference between two results that the node with the maximum centrality value for the latter case is more close to the network center. The reason is that the flow betweenness is based on the paths between nodes, while the approximation flow betweenness is concerned in the ability that one node collects the information around. Although there is a little error between the approximation method and the original method, the approximation result can also reflect the characteristics of the network very well, and thus we can save a lot of computation time. If a network has N nodes, the time complexity is about $O(N^3)$ for the flow betweenness, while $O(N^2)$ for the approximation measure.

In the process of information flowing, since the flowing direction is free, a few factors may affect the result. The severest factor is that some information may just be flowed between two or several nodes. In this case, the network can be pretreated before evaluation. For example, in the example network as shown in Fig. 3, Node 2 and Node 8 can be combined into one node that should be endowed a new value, which is the sum of the values appointed to the two nodes. Based on the same rule, Node 5, Node 6 and Node 13 can also be combined into one node. After this pretreatment, the result will be more precise.

3. Example

This section gives an example. The network in Fig. 4 de-

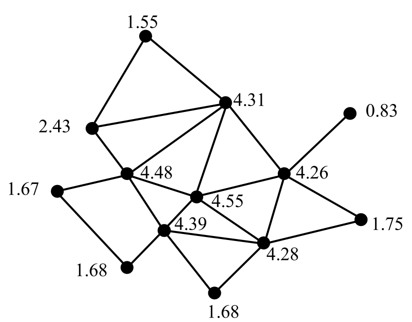

Fig. 2 Results of the approximate calculation.

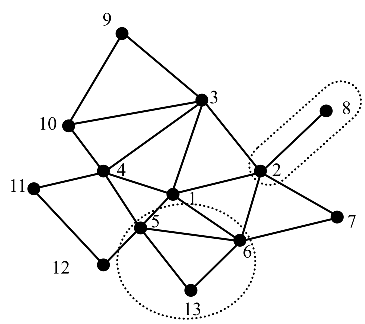

Fig. 3 Pretreatment before the approximate calculation.

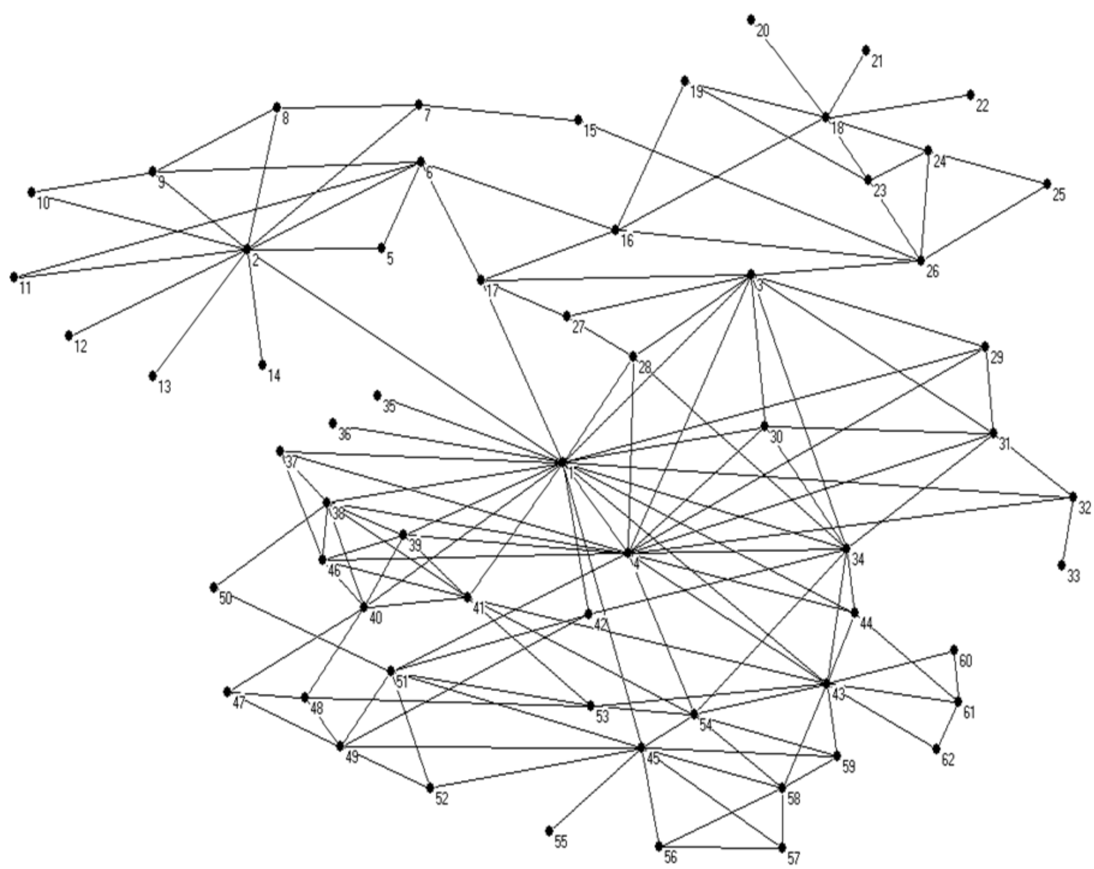

Fig. 4 The structure of the test network.

Table 1 The results of the approximate calculation scheme.

Node	1	4	2	43	3
Result	17.55	13.31	10.60	10.13	8.71
Node	45	34	54	6	18
Result	8.54	8.51	7.01	6.95	6.93
Node	41	26	38	16	51
Result	6.71	6.56	6.46	6.20	6.16
Node	40	49	58	17	31
Result	5.77	5.20	5.16	5.15	5.05

picts an air kidnapping incident that happened in New York in November 2001. This network is a random network whose degree distribution is exponential. In this network, nodes represent the terrorists, and edges represent their contact. We adopt the approximate calculation method to evaluate the importance of each node, obtaining the approximate flow betweenness values for all nodes. We list the 20 largest values in Table 1 in the descending order. It is easy to find that Node 1 has the biggest value, and it is most important node in the network. Node 4, Node 2, Node 43 and Node 3 are with the following bigger values, thus they are also important nodes in this network. In the example network, all above nodes close to the network center. In a word, the approximate calculation works well in finding important nodes in a network.

4. Conclusions

In this Letter, we have proposed an approximate calculation method for the flow betweenness centrality. By calculating the ability that a node collects the information around it, we obtain an approximate result of this node's flow betweenness. Then we give a simple comparison between two methods and also provide the experimental results as a proof. The time complexity of our approach is $O(N^2)$, which is far less than that of the original flow betweenness. Since our approach avoids calculating the shortest path between nodes, it is more efficient than many existing approaches for betweenness calculation. However, the proposed method can only be used in the static network with complete information. We have not verified the accuracy under the situations that the network information is incomplete or the network is dynamic. Future work will concentrate on obtaining a general formula that can be used for different conditions.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under grant 61171150 and 61070208, and also supported by Zhejiang Provincial Natural Science Foundation of China under grant R1110006.

References

[1] D.J. Watts and S.H. Strogatz, "Collective dynamics of 'small-world' networks," Nature, vol.393, pp.440–442, June 1998.
[2] M.E.J. Newman and D.J. Watts, "Renormalization group analysis of the small-world network model," Phys. Lett. A, vol.263, pp.341–346, Dec. 1999.
[3] A.L. Barabási and R. Albert, "Emergence of scaling in random networks," Science, vol.286, pp.509–512, Oct. 1999.
[4] E. Costenbade and T.W. Valente, "The stability of centrality measures when networks are sampled," Social Networks, vol.25, pp.283–307, Oct. 2003.
[5] L.C. Freeman, "Centrality in social networks conceptual clarification," Social Networks, vol.1, pp.215–239, 1979.
[6] H.W. Ma and A.P. Zeng, "The connectivity structure, giant strong component and centrality of metabolic networks," Bioinformatics, vol.19, pp.1423–1430, Feb. 2003.
[7] L.C. Freeman, D. Roeder, and R.R. Mulholland, "Centrality in social networks: II. experimental results," Social Network, vol.2, pp.119–141, 1979.
[8] L.C. Freeman, S.P. Borgatti, and D.R. White, "Centrality in valued graphs: A measure of betweenness based on network flow," Social Networks, vol.13, pp.141–154, June 1991.
[9] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," Social Networks, vol.23, pp.191–201, July 2001.
[10] E. Estrada and J. Rodríguez-Velázquez, "Subgraph centrality in complex networks," Phys. Rev. E, pp.056103-1–9, May 2005.
[11] M.E.J. Newman, "A measure of betweenness centrality based on random walks," Social Networks, vol.294, pp.39–54, Jan. 2005.
[12] U. Brandes, "A faster algorithm for betweenness centrality," J. Mathematical Sociology, vol.25, pp.163–177, April 2001.
[13] D.A. Bader, S. Kintali, K. Madduri, and M. Mihail, "Approximating betweenness centrality," Proc. 5th Workshop on Algorithms and Models for the Web-Graph, LNCS 4863, pp.124–137, San Diego, CA, Dec. 2007.
[14] R. Geisberger, P. Sanders, and D. Schultes, "Better approximation of betweenness centrality," Proc. Ninth Workshop on Algorithm Engineering and Experiments, pp.90–100, San Francisco, California, USA, Jan. 2008.
[15] D. Gkorou, J. Pouwelse, and D. Epema, "Efficient approximate computation of betweenness centrality," Proc. 16th Annual Conf. Advanced School for Computing and Imaging, Veldhoven, Netherlands, Nov. 2010.
[16] M.J. Lee, J. Lee, J.Y. Park, R.H. Choi, and C.W. Chung, "QUBE: A quick algorithm for updating betweenness centrality," Proc. 21st International Conference on World Wide Web, pp.351–360, Lyon, France, April 2012.