Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

International Journal of Modern Physics C
Vol. 29, No. 12 (2018) 1850125 (11 pages)
© World Scientific Publishing Company
DOI: 10.1142/S0129183118501255

World Scientific
www.worldscientific.com

Evaluation method for node importance based on attraction between nodes

Jin Zeng*, Chenxi Shao†, Xingfu Wang‡ and Fuyou Miao§

University of Science and Technology of China
Hefei, P. R. China
**zengjin@mail.ustc.edu.cn*
†cxshao@ustc.edu.cn
‡Wangxfu@ustc.edu.cn
§mfy@ustc.edu.cn

Received 12 April 2018
Accepted 5 December 2018
Published 7 January 2019

Vital node, which has some special functions, plays an important role compared to other nodes in complex networks. Recently, the discovery of vital nodes in complex networks has captured increasing attention due to their important theoretical significance and great practicability. By defining the confidence of the node and the inter-node attraction, the significance of the node is measured by the product of the confidence of the node and the aggregation of attractions of the node on other nodes in the network. The experimental results illustrate that the proposed method has higher precision and performs well on various networks with different structures.

Keywords: Complex network; vital node; inter-node attraction.

PACS Nos.: 11.25.Hf, 123.1K.

1. Introduction

Complex networks are abstractions of complex systems. Social network, the World Wide Web, citation network, transportation network, etc., are all abstracted from complex systems. Previous researches demonstrate that networks with different topologies have different invulnerabilities against different type of attacks.[1] For example, in random attacks, the scale-free networks have better fault-tolerance than random networks, and attack certain nodes deliberately, it is very vulnerable. In other words, the distribution of key nodes is different in networks with different topologies. The importance gap between the nodes in random network is not obvious, and in scale-free network, the importances of nodes are polarized. Identifying the key nodes in the network and protecting them can improve the reliability of the entire network. Many indicators for measuring the importance of the nodes have been

† Corresponding author.

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

proposed.[2] These indicators have their own advantages as well as disadvantages, such as the degree centrality has a low time complexity but poor performance, the betweenness centrality performs well but costs $O(n^3)$ time. In addition to these indicators, some researchers have proposed some new evaluation methods for measuring node importance: First, the cohesion of the network has been defined, the importance of a node is measured by the increment of cohesion of the network after the node shrinks.[6] It is proposed to measure the importance of a node by the number of minimum spanning trees after deleting the node in Ref. 8, etc. However, they are not impeccable, as the method proposed in Ref. 6, they cannot distinguish the shrunk nodes due to the same topologies of the network. Considering that most existing methods are not sufficiently precise or the time complexity for computation is high, a new evaluation method for node importance based on inter-node attraction is proposed in this paper (MBA). First, the confidence of the node and the inter-node attraction have been defined, then the importance of a node is measured by the product of the value of the confidence of the node and the sum of attractions the node holds for other nodes in the network. Compared with most existing methods, we consider both the locally information of the node and the globally information of the network, so the results obtained by the method are more accurate. The method has been applied to various real-world networks and artificial networks, compared with other methods mentioned above, the proposed method is more accurate and able to adapt to various networks with different topologies.

2. Method

Intuitively, the importance of the node can be reflected by its influence on the network. However, it is not easy to define the influence of the node on the network, since the network is composed of connected nodes. Therefore, we evaluate the influence of the node on the network by the sum of the attractions the node holds for other nodes in the network. The key point is how to calculate the inter-node attraction. In consideration of the fact that the objects with same properties tend to have higher affinity, i.e. water and alcohol can blend but not with oil; individuals with more common interests and hobbies are more likely to be good friends. Therefore, isomorphism between nodes was defined, and the attraction between nodes is positively related to the isomorphism. The Isomorphism $H(i,j)$ and the attraction $f(i,j)$ are defined as

$$H(i,j) = \frac{\sum_{k=1}^{n} A_{ik}A_{jk} + 1}{\left((\sum_{k=1}^{n} A_{ik}^2 + 1)(\sum_{k=1}^{n} A_{jk}^2 + 1)\right)^{\frac{\theta}{2}}}, \quad (1)$$

$$f(i,j) = \frac{H(i,j)}{d^2}, \quad (2)$$

where A is adjacency matrix of the network, isomorphism $H(i,j)$ is positively related to the number of common neighbors and negatively correlated with the product of

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

the degrees of two nodes. Since most nodes are not sharing any neighbor node, we construct a virtual node that connects with each node in the network. The attraction $f(i, j)$ between nodes is positively correlated with $H(i, j)$ and negatively correlated with the square of the shortest distance d between these two nodes. The parameter θ in Eq. (1) controls the influence of the degree of the node on isomorphism $H(i, j)$. For example, the degree of node has a great influence on the isomorphism when the parameter θ is large, under this scenario, the bridge nodes in networks always tend to be more important. Certainly, the importance of bridge nodes should be emphasized but also limited. Therefore, it is necessary to find a suitable value of θ.

With the formula for calculating the attraction between nodes, the attraction of a node for the rest of the nodes in the network can be calculated and then the nodes can be ranked in the network by the index. However, there is a flaw in this scheme, that is, when the attraction of a node for few of the nodes is large, but for others is weak, the node is still considered to be important because of the support of few of the nodes. As a consequence, it brings difficulty in finding the globally important nodes. In order to solve this problem, the supporter and the confidence C_i of a node are defined. The supporter of a node is defined as the node where the attraction between them exceeds the average, the support rate P_i of the node is the ratio of the number of supporters to the total number of nodes in the network. The confidence of the node is positively related to the value of its support rate. Then, the importance of a node is measured by the product of its confidence and attraction for other nodes in the network, in this way, it can avoid the case that the locally important nodes become globally important nodes because of the support of a few of the nodes.

Since the support rate of each node is between 0 and 1, in order to make the confidence C_i change smoothly with the support rate P_i so that the importance of a node can mainly depend on the attraction, the confidence C_i is defined as

$$C_i = e^{P_i}. \tag{3}$$

According to the above, the importance I_i of a node is defined as

$$I_i = C_i \sum_{j=1}^{n} f(i, j). \tag{4}$$

The importance of each node in the network could be calculated according to Eq. (4). Then, we rank these nodes based on their importance as shown in the following algorithm:

(1) Calculate the shortest distance between each node pair by Floyd algorithm to get the shortest distance matrix.
(2) Select each node in turn and calculate the sum of attraction for the remaining nodes.
(3) Calculate the confidence C_i and importance I_i of the node.
(4) Check whether traversed all nodes in the network, if not, goto step 2.
(5) Rank all nodes according to their importance I_i.

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

The time complexity of this algorithm depends on the calculation of the shortest distance between pairs of nodes, because the time complexity of Floyd algorithm is $O(n^3)$, our algorithm also requires $O(n^3)$ time. However, because of the small-world nature of complex networks,[11,12] Floyd algorithm could be optimized for this feature and the time complexity reduces to $O(Dn^2)$, where D is the diameter of the network.[11]

3. Experiment

Figure 1 shows that there are eight nodes and eight edges, obviously, the importances of nodes cannot be distinguished by the degree centrality. For example, the degrees of node 3, node 5, node 6 are both 3, but their significances should be different, because two of the three nodes adjacent to node 3 are edge nodes, and the node 5 and node 6 are different from node 3, e.g. although the numbers of neighbor nodes of the three nodes are the same, the quality is different. As for node 4, although its degree is 2, it is a bridge node, its importance should be greater. The method proposed in Ref. 6 cannot distinguish between the node 5 and the node 6, because the network topology obtained by shrinking node 5 or node 6 is the same. Although the betweenness centrality has good discrimination ability, it is still powerless for some special node. For example, the betweenness centralities of the node 4 and node 5 in Fig. 1 are the same, but their degree centralities are different. Therefore, they should be considered not equally important. The reason that the degree centrality, the betweenness centrality and the method proposed in Ref. 6 are insufficiently accurate is that these methods consider only one of the globally properties or locally properties of the nodes but not all. The method proposed in this paper takes the number of the

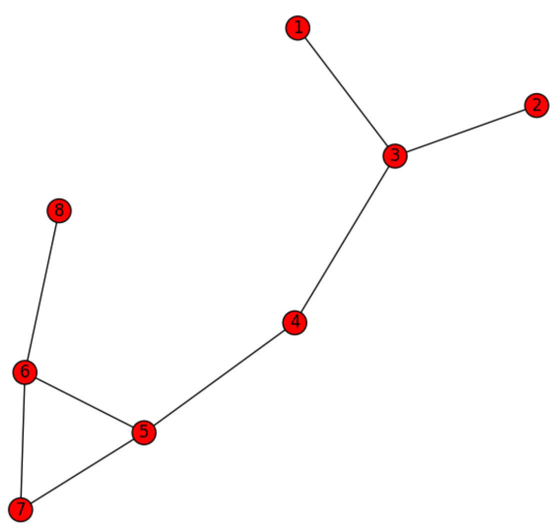

Fig. 1. (Color online) A simple artificial network with 8 nodes.

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

Table 1. The evaluation results of the simple network obtained by different methods (the higher the score is, the more important the node is).

Node	BC	Literature[6]	DC	MBA ($\theta = 0.6$)	MBA ($\theta = 2.0$)
V1	0.00	0.19	1	1.91	0.55
V2	0.00	0.19	1	1.91	0.55
V3	0.52	0.59	3	2.76	0.57
V4	0.57	0.47	2	3.45	0.74
V5	0.57	0.54	3	4.22	0.82
V6	0.28	0.54	3	3.94	0.70
V7	0.00	0.35	2	3.31	0.73
V8	0.00	0.21	1	1.81	0.46

supporters of a node as the locally property of the node and the sum of attractions for the other nodes in the network as the globally property, so it performs better on evaluating the importances of nodes in networks. The experimental result is shown in Table 1.

As Table 1 shows, the method proposed in this paper achieved better result and adapted to different requirements, i.e. the greater the value of θ, the greater the importance of bridge nodes. As the table shows, when $\theta = 0.6$, the node 4 ranks third, but when $\theta = 2.0$, the node 4 exceeds the node 6, rising to second place. In order to further illustrate the effectiveness of the method, we apply it on karate network, arpa network, and various artificial networks and the SIR model as the benchmark is used to evaluate the efficiency of our proposed method through examining the spreading influence of top-k ranked nodes. For simplicity, we set the infection rate $\alpha = 0.1$ and the recovery rate $\beta = 1.0$. Initially, there is only one infected node, we agree that the node which eventually infects more the other nodes is more important when the infection process ends. The network topology of karate network is shown as Fig. 2,

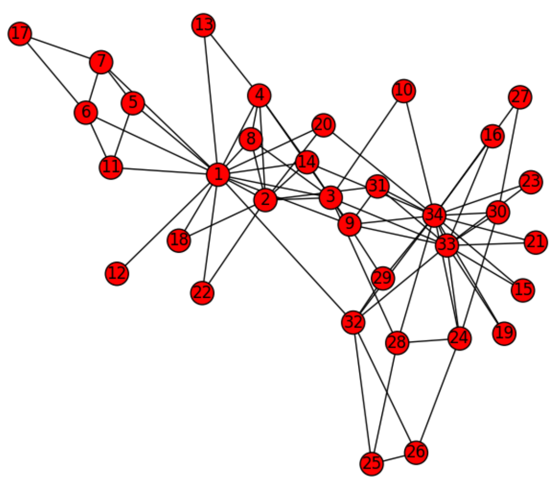

Fig. 2. (Color online) Visualization of the social relationships among the 34 individuals in the karate club.

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

Table 2. The top-10 nodes in karate network picked by different methods.

Ranking	DC	BC	Lit[6]	MBA ($\theta = 0.6$)	SIR ($\alpha = 0.1$, $\beta = 1.0$)
1	34	1	34	1	1
2	1	34	1	34	34
3	33	33	33	33	33
4	3	3	3	3	3
5	2	32	32	2	2
6	4	9	2	32	9
7	32	2	14	14	14
8	9	14	9	4	4
9	14	20	4	9	32
10	24	6	20	31	31

the network was studied by Zachary for a period of three years from 1970 to 1972, the network captures 34 members of a karate club, documenting 78 pairwise links between members who interacted outside the club, and the node 1 represents the coach of the club, the node 34 represents the president of the school, the club was divided into two small groups because of conflict between them.

The performances by using different methods on karate network are shown in Table 2.

As Table 2 shows, our method (MBA) is most similar to the average result obtained by repeating tests 10 000 times on karate network using the SIR model. Besides, both of them consider the node 31 as one of the top-10 nodes, while the node 31 is not selected among the top-10 nodes using the other three methods. In fact, the node 31 surrounded by the small group leads by node 1 and the small group leads by 34 node, such a bridge node should be somewhat important. Arpa network captures 21 nodes and 23 edges, the average degree of nodes in the network is between 2 and 3, so the degree centrality is ineffective. It is a common network for which we test the efficiency of evaluation method for node importance, the topology of Arpa network is shown as Fig. 3.

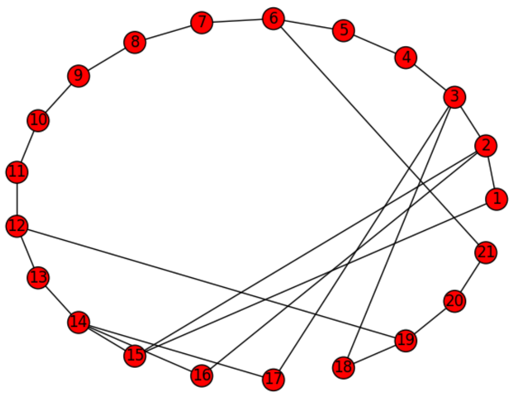

Fig. 3. (Color online) Advanced research projects agency network (Arpa network).

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

Table 3. The rankings obtained by different methods and the number in parentheses is the score of the node.

Ranking	BC	Lit[6]	MBA ($\theta = 0.6$)	SIR ($\alpha = 0.1$, $\beta = 1.0$)
1	3(0.316)	3(0.308)	2(5.54)	2
2	12(0.271)	14(0.275)	3(5.31)	3
3	19(0.215)	12(0.261)	15(4.95)	14
4	6(0.201)	6(0.255)	14(4.38)	15
5	4(0.179)	2(0.251)	17(4.08)	12
6	14(0.175)	19(0.231)	19(3.90)	19
7	13(0.159)	4(0.191)	12(3.88)	17
8	5(0.157)	5(0.191)	18(3.79)	16
9	11(0.156)	13(0.191)	1(3.78)	6
10	2(0.149)	15(0.186)	13(3.72)	13
11	18(0.120)	7(0.183)	16(3.65)	1
12	10(0.107)	8(0.183)	6(3.37)	18
13	7(0.106)	9(0.183)	4(3.33)	4
14	20(0.097)	10(0.183)	20(2.88)	20
15	21(0.072)	11(0.183)	5(2.87)	7
16	9(0.072)	18(0.167)	11(2.85)	5
17	8(0.069)	20(0.150)	21(2.81)	10
18	17(0.047)	21(0.150)	7(2.76)	21
19	15(0.046)	17(0.149)	10(2.40)	11
20	16(0.012)	1(0.127)	8(2.35)	9
21	1(0.00)	16(0.125)	9(2.34)	8

The performances by using different methods on Arpa network are shown in Table 3.

The experimental result using our proposed method is comparable with the average result obtained by SIR model and different from the other two methods. The reason is that the betweenness centrality and the method proposed in Ref. 6 focus on the contribution of the node to the shortest routes of the network and ignore the suboptimal routes. However, the principle of the proposed method is similar to the SIR model considering that the importance of a node is positively related to the influence of the node on other nodes in the network and the influence is not spread only through the shortest path. There are varying degrees of insufficient precision in the betweenness centrality and the method proposed in Ref. 6. For example, the importances of node 9 and node 21 cannot be figured out by the betweenness centrality. The method proposed in Ref. 6 is incapable of figuring out the importances of node 7, node 8, node 9, node 10 and node 11. The method proposed in the paper has a higher precision, because the factors considered are more comprehensive. The last real-world network is the railway network of China, the topology of the network is shown in Fig. 4. We apply each method on the network, the results are shown in Table 4.

As Table 4 shows, our method considers Wuhan station is most important. Just consider the location, Wuhan is the thoroughfare to nine provinces. Although it is a second-tier city, the passenger traffic of Wuhan station ranks first in China.

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

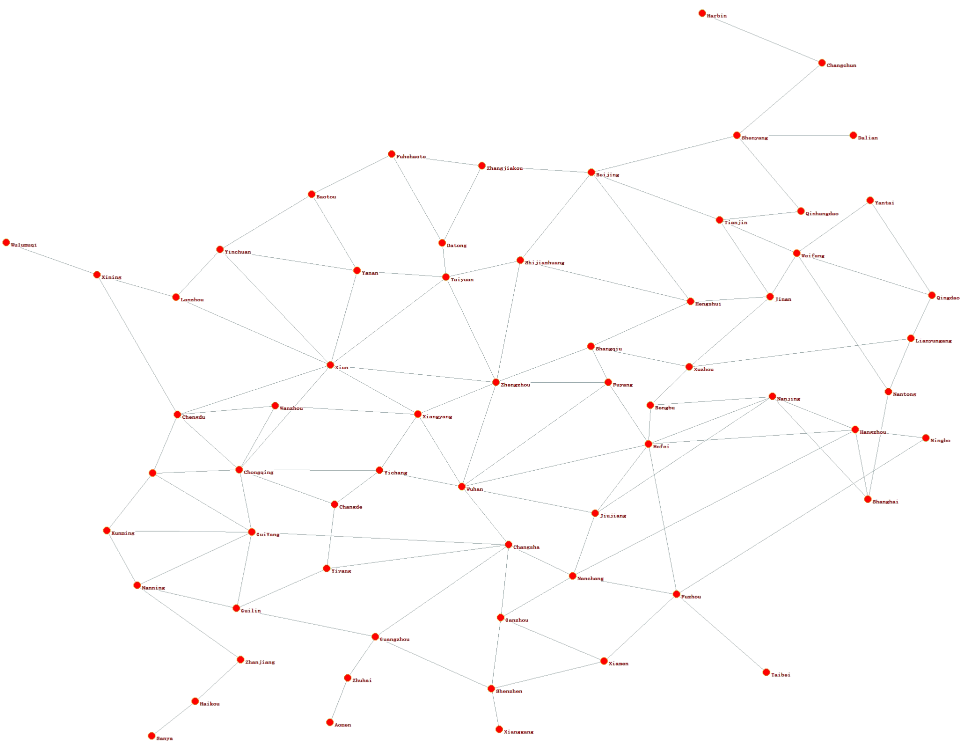

Fig. 4. (Color online) The topology of the railway network of China.

Table 4. The top-5 railway station selected by different methods.

Ranking	BC	H-index	CC	MBA
1	Zhengzhou station	Xian station	Zhengzhou station	Wuhan station
2	Changsha station	Chongqing station	Wuhan station	Zhengzhou station
3	Wuhan station	Xiangyang station	Xiangyang station	Xian station
4	Xian station	Zhengzhou station	Xian station	Chongqing station
5	BeiJing station	Taiyuan station	Fuyang station	HeFei station

In order to illustrate that this algorithm can be applied to networks with different topologies, the method and some other methods were applied to three simulation networks with different topologies. We picked the top-10 nodes from thousands of nodes as a set, and used the average result obtained by repeating SIR test 10 000 times as a benchmark, taking the Jaccard similarity coefficient of the two sets as a measurement of the effectiveness of the method. The experimental results are shown in Figs. 5–7. As the figures show, on small-world network and ER-network, the proposed method has obvious advantage over the other methods. Especially on small-world network, the results obtained by the other methods are quite different from those obtained by the SIR model. The reason that these methods underperform

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

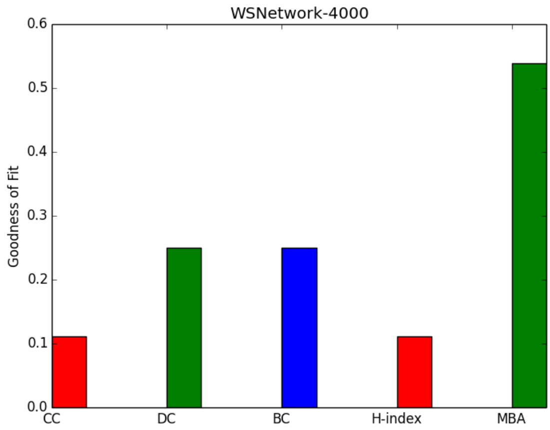

Fig. 5. (Color online) The jaccard coefficients of the sets obtained by various methods on the small-world network which captures 4000 nodes.

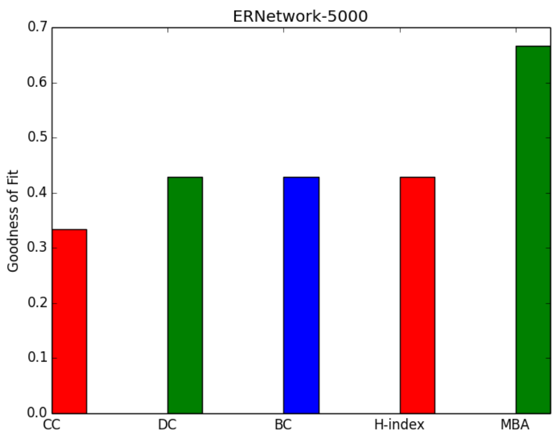

Fig. 6. (Color online) The jaccard coefficients of the sets obtained by various methods on the ER-network which captures 5000 nodes.

is that they are not sufficiently precise, i.e. there are many nodes that cannot be distinguished by these methods in small-world network because of the special topology. These methods all perform well on BA scale-free network, which is also determined by the characteristics of the scale-free network. In the formation process of a scale-free network, the new node always tends to connect to a node that has many neighbors. For example, a new paper is more likely to cite highly cited papers in this field in citation network. The characteristic of scale-free network causes the polarization of the importances of nodes. Therefore, almost all evaluation methods for node importance can perform well on the scale-free networks.

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

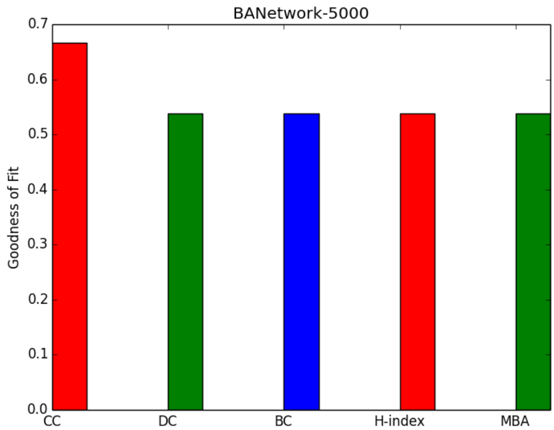

Fig. 7. (Color online) The jaccard coefficients of the sets obtained by various methods on the BA scale-free network which captures 5000 nodes.

4. Conclusion and Discussion

Aiming at the problem that most existing methods are not sufficiently precise, a method that takes account of the location and the number of supporters of the node has been proposed. The experimental results show that our method provides better results in various conditions (the sizes of networks, the topologies of networks). However, the parameter θ in Eq. (1) has still to be discussed, the values of θ used in the experiments may not be the optimal, so how does one find appropriate value of θ to satisfy the various requirements? We consider it as the promising direction for future work.

Acknowledgments

The work is supported by the NNSFC Nos. 61472381, 61174144, 61472382, 61772490 and 61572454.

References

1. H. Paul and B. Seth, *2008 Proc. of the 41st Annual Hawaii International Conf. System Sciences* (Hawaii, January, 2008), p. 1.
2. L. L. Chen *et al.*, *Phys. Rep.* **650**, 1 (2016).
3. A. Szolnoki *et al.*, *Phys. Rev. E Statistic. Nonlinear Soft Matter Phys.* **87**, 042805 (2013).
4. M. E. J. Newman, *Siam Rev.* **45**, 167 (2006).
5. J. M. Kleinberg, *J. ACM* **46**, 604 (1999).
6. T. Yuejin, W. Jun and D. Hongzhong, *Syst. Eng. Theor. Pract.* **26**, 79 (2006).
7. S. Brin, *Proc. 7 Int. World Wide Web Conf.* **56**, 3825 (1998).
8. Y. Chen, A. Q. Hu and X. Hu, *J. Chin. Institute Comm.* **25**, 129 (2004).
9. M. Kitsak *et al.*, *Nat. Phys.* **6**, 888 (2010).
10. L. Lv *et al.*, *Nat. Comm.* **7**, 10168 (2016).
11. X. Zhou *et al.*, *Acta Physica Sinica* **61**, 485 (2012).

Int. J. Mod. Phys. C 2018.29. Downloaded from www.worldscientific.com
by UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA on 04/29/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

12. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
13. A. Barabsi and R. Albert, *Science* **286**, 509 (1999).
14. L. Zhong *et al.*, Identifying influential nodes in complex networks: A multiple attributes fusion method, *Int. Conf. Active Media Technology* (Springer, Cham, 2014), pp. 11–22.
15. D. Chen *et al.*, *Phys. A Statistic. Mech. Appl.* **391**, 1777 (2012).
16. J. G. Liu *et al.*, *Acta Physica Sinica* **62**, 178901 (2013).