Identifying influential spreaders in complex networks based on entropy weight method and gravity law*

Xiao-Li Yan(闫小丽)[1,2,3], Ya-Peng Cui(崔亚鹏)[1,2,3], and Shun-Jiang Ni(倪顺江)[1,2,3,†]

[1] *Institute of Public Safety Research, Tsinghua University, Beijing* 100084, *China*
[2] *Department of Engineering Physics, Tsinghua University, Beijing* 100084, *China*
[3] *Beijing Key Laboratory of City Integrated Emergency Response Science, Beijing* 100084, *China*

(Received 21 November 2019; revised manuscript received 30 January 2020; accepted manuscript online 20 February 2020)

In complex networks, identifying influential spreader is of great significance for improving the reliability of networks and ensuring the safe and effective operation of networks. Nowadays, it is widely used in power networks, aviation networks, computer networks, and social networks, and so on. Traditional centrality methods mainly include degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, k-shell, *etc*. However, single centrality method is one-sided and inaccurate, and sometimes many nodes have the same centrality value, namely the same ranking result, which makes it difficult to distinguish between nodes. According to several classical methods of identifying influential nodes, in this paper we propose a novel method that is more full-scaled and universally applicable. Taken into account in this method are several aspects of node's properties, including local topological characteristics, central location of nodes, propagation characteristics, and properties of neighbor nodes. In view of the idea of the multi-attribute decision-making, we regard the basic centrality method as node's attribute and use the entropy weight method to weigh different attributes, and obtain node's combined centrality. Then, the combined centrality is applied to the gravity law to comprehensively identify influential nodes in networks. Finally, the classical susceptible-infected-recovered (SIR) model is used to simulate the epidemic spreading in six real-society networks. Our proposed method not only considers the four topological properties of nodes, but also emphasizes the influence of neighbor nodes from the aspect of gravity. It is proved that the new method can effectively overcome the disadvantages of single centrality method and increase the accuracy of identifying influential nodes, which is of great significance for monitoring and controlling the complex networks.

Keywords: complex networks, influential nodes, entropy weight method, gravity law

PACS: 89.75.Hc **DOI:** 10.1088/1674-1056/ab77fe

1. Introduction

With the development of science and technology and the progress of society, various networks are gradually formed, such as aviation networks,[1] traffic networks,[2] computer networks,[3] social networks,[4] and biological networks,[5] which are closely related to our life and work. In recent decades, the researches on complex networks have gradually attracted the attention of scholars from all walks of life. It is found that in many real-society networks, different nodes have different effects on the networks.[6–8] Therefore, identifying influential nodes is of great significance for stably operating and effectively controlling the networks, such as finding social leader,[9] mitigating disease spreading,[10] controlling information dissemination,[11,12] detecting community structures,[13] *etc*., and it has become one of the research hotspots of complex networks.

In the past researches, scientists have proposed a series of centrality methods, such as the degree centrality (DC),[14,15] betweenness centrality (BC),[16,17] closeness centrality (CC),[16] eigenvector centrality (EC),[18] k-shell,[7] PageRank (PR),[19] H-index algorithm,[19] *etc*. However, sometimes these centrality methods are one-sided and inaccurate in identifying influential nodes in complex networks. For example, degree centrality only considers the local information about nodes, regardless of the role of neighbor nodes; the CC and BC focus on the shortest path, but information is not always transmitted through the shortest path, and the time complexity is high; when the degree of nodes is large, the EC has the phenomenon of local centralization of numerical value; The k-shell and H-index algorithm cannot distinguish between nodes with the same centrality value; PR ignores the situation of the page itself, and does not distinguish between the types of links in the page.

In order to improve the accuracy of identifying influential nodes, researchers have also proposed a great many of improved centrality methods. For instance, Wen and Deng[20] proposed a new method based on the local information dimensionality (LID) of nodes, in which the quasi-local information around each node is considered. Fei *et al.*[21] proposed a novel method based on the inverse-square law by defining the intensity of node. Gao *et al.*[22] proposed a local structural centrality measuring method by considering the number

*Project support by the National Key Research and Development Program of China (Grant No. 2018YFF0301000) and the National Natural Science Foundation of China (Grant Nos. 71673161 and 71790613).

†Corresponding author. E-mail: sjni@tsinghua.edu.cn

of nearest neighbors of a node and the topological connections among the neighbors. By taking into account the neighbors' resource and the influence of spreading rate for the target node, Zhong *et al.*[23] proposed an improved iterative resource allocation (IIRA) method to identify influential nodes. Wang *et al.*[24] combined degree and weight strength of each node and proposed a modified efficiency centrality (EffCm) in weighted network. Zeng and Zhang[25] proposed a mixed degree decomposition (MDD) method by incorporating the residual degree and the exhausted degree to revise the original k-shell decomposition, but it is difficult to find the optimal parameter λ to achieve better results. Instead of ranking all nodes, Song *et al.*[26] aimed at ranking a small number of nodes in networks and proposed a rapid identifying method (RIM) to find the fraction of high-influential nodes.

Recently, based on the multiple attribute decision making (MADM),[14,27–30] Mo and Deng[31] proposed a multi-evidence centrality (MeC) method from the perspective of multi-attributes. Bian *et al.*[32] considered several different centrality methods as the multiple attributes and proposed a method to identify influential nodes based on the analytic hierarchy process (AHP). Hu *et al.*[33] proposed a weighted technique to identify influential nodes based on the technique for order preference by similarity to an ideal solution (TOPSIS) by calculating the weight of each attribute. To further explain the influence of neighbor nodes, Li *et al.*[34] proposed a gravity model in which both neighborhood information and path information are used to measure a node's influence. By viewing the k-shell value of each node as its mass and the shortest path distance between two nodes as their distance, according to the gravity formula, Ma *et al.*[6] proposed a gravity centrality index to identify the influential spreaders.

Based on the above researches, it can be found that a common shortcoming in these centrality methods is that in most of methods only single centrality characteristic is considered, or the weight of every characteristic is regarded as being identical when multiple characteristics are considered, which may reduce the accuracy of ranking nodes. Another obvious shortcoming is that the influence of a node depends not only on its direct neighbors (1-step neighbors), but also on its 2-step and even more steps of neighbors. So to solve these issues, we need to do further research on identifying influential nodes. In view of the multi-attribute decision-making, we propose an improved centrality method in which taken into account are several aspects of node's properties, including local topological characteristics, central location of nodes, propagation characteristics, and properties of neighbor nodes. We regard the classical centrality methods as the attributes of nodes. Considering the different effects of different attributes on nodes, we use entropy weight method,[14,28,29] which is a method to calculate the weight quantitatively, to weigh different attributes and obtain the combined centrality. Then the gravity law[6,34,35] is utilized to evaluate the influence of nodes. Finally, based on the real complex networks, the classical SIR[36] model is used to simulate the spreading process and verity the evaluation results. The results show that the proposed method is feasible and accurate in identifying influential nodes.

With the increasing variety and quantity of complex networks in society, there is an urgent need for some efficient, accurate and universally applicable centrality methods. Based on our proposed method, we can find the influential nodes in the networks, focus on them, protect and regulate them, so as to improve the reliability of the network. For example, in the transmission of infectious diseases, we can determine "super communicators", predict the trend and scope of virus transmission, and provide a basis for virus prevention and control; in the traffic network, we can evaluate the traffic congestion to further ensure the connectivity of the network based on influential nodes. And the method can be applied to all kinds of networks, whether it is directed or undirected, weighted or unweighted. Therefore, nowadays, our research is of great significance in complex networks.

The rest of the paper is organized as follows. In Section 2, the description of our method is presented. In Section 3, we introduce the datasets, the spreading model and the evaluation methodologies that are used to evaluate the performance of our method. In Section 4, we analyze the effectiveness of our proposed method in six real networks through the numerical simulation of classical SIR model. And finally, some conclusions are drawn from the present study and brief discussion is also presented in Section 5.

2. Method

As stated above, a lot of researches have been carried out to identify influential nodes in complex networks and many centrality methods have also been proposed. These centrality methods show their advantages and superiorities together with some shortcomings and limitations. Based on these researches, in this paper we utilize the idea of the multi-attribute decision-making and propose a novel centrality method of identifying the influential nodes in the complex networks based on entropy weight method and gravity law.

Firstly, we summarize four important characteristics relating to the influence of node, namely local topological characteristic, central location of nodes, propagation characteristics and properties of neighbors. Then we regard seven classical centrality methods selected as the attributes of nodes, and according to the expression meanings of different centrality methods, we classify seven attributes as four categories.

Table 1. Four categories of node attributes.

Category	1 local topological characteristic	2 central location of nodes	3 propagation characteristic	4 properties of neighbors
Centrality (attribute)	DC H-index	CC k-shell	BC	EC PageRank

Secondly, based on the idea of the multi-attribute decision-making, we select different attributes to form different combinations according to Table 1. Each combination contains four attributes selected from four categories respectively. So eight combinations are obtained. Besides, the combination of all seven attributes is regarded as a contrast. The specific combinations are as follows.

Combination 1: DC, CC, BC, EC;

Combination 2: DC, CC, BC, PageRank;

Combination 3: DC, k-shell, BC, EC;

Combination 4: DC, k-shell, BC, PageRank;

Combination 5: H-index, CC, BC, EC;

Combination 6: H-index, CC, BC, PageRank;

Combination 7: H-index, k-shell, BC, EC;

Combination 8: H-index, k-shell, BC, PageRank;

Combination 9: DC, H-index, CC, k-shell, BC, EC, PageRank.

Then, because the influences of different attributes on nodes are often different, we use the entropy weight method to calculate the weight of each attribute in each combination quantitatively. Generally speaking, the smaller the information entropy of an index, the larger the variation of the index is and the larger its weight. By weighted averaging, we obtain the combined centrality of each node in each combination. The specific calculating steps are as follows.

Assuming that there are N nodes in the network, the set of research objects can be expressed as $X = \{x_1, x_2, \ldots, x_N\}$. If there are m centrality methods, the set of attributes of nodes can be expressed as $A = \{a_1, a_2, \ldots, a_m\}$.

Step 1 Constructe the decision matrix of all nodes and attributes in complex networks $\boldsymbol{P} = (x_{ij})_{N\times m}$

$$\boldsymbol{P} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nm} \end{bmatrix}. \tag{1}$$

In the decision matrix $\boldsymbol{P}$, x_{ij} $(i = 1,2,\ldots,N;\ j = 1,2,\ldots,m)$ represents the j-th attribute of the i-th node.

Step 2 Normalize decision matrix to obtain normalized decision matrix $\boldsymbol{B} = (b_{ij})_{N\times m}$

$$b_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}}. \tag{2}$$

$x_j^{\max} = \max\{x_{ij} | 1 \leqslant i \leqslant N\}$; $x_j^{\min} = \min\{x_{ij} | 1 \leqslant i \leqslant N\}$.

Step 3 Calculate the information entropy for each attribute

$$S_j = -K \sum_{i=1}^{N} P_{ij} \ln P_{ij}. \tag{3}$$

In the formula, $j = 1,2,\ldots,m$, $K = 1/\ln N$; $P_{ij} = b_{ij} / \sum_{j=1}^{m} b_{ij}$, if $P_{ij} = 1$, $P_{ij} = (1+b_{ij})/(\sum_{j=1}^{m}(1+b_{ij}))$.

Step 4 Calculate the entropy weight for each attribute.

The set of the information entropy of each index, $[S_1 S_2 \ldots S_m]$, is as follows:

$$\omega_j = \frac{1 - S_j}{m - \sum_{j=1}^{m} S_j}. \tag{4}$$

Step 5 Calculate the combined centrality of each node

$$C_i = S_1 \cdot b_{i1} + S_2 \cdot b_{i2} + \ldots + S_m \cdot b_{im}. \tag{5}$$

Ultimately, based on the gravity law first proposed by Newton, a physicist, we replace the mass with the combined centrality of nodes, and regard the shortest path distance between target node and its neighbor in network as their distance. In this way, the influence of node i in our proposed method is denoted as $G(i)$ and expressed as follows:

$$G(i) = \sum_{j \in \psi_i} \frac{C_i C_j}{d_{ij}^2}. \tag{6}$$

In the formula, ψ_i is the neighborhood set whose distance to node i is less than or equal to a given value r, d_{ij} is the shortest path distance between node i and node j, C_i, and C_j are the combined centrality of the target node i and its neighbor node j, respectively. In our research, we take $r = 3$, namely, only the nearest neighbor, the second nearest one and the third nearest one are considered.

On the whole, our method has several obvious advantages. Compared with other centrality methods, such as DC, BC, and CC, firstly, in our method more attention is paid to the multiple characteristics of nodes. Secondly, the attributes of nodes are not combined randomly. According to the four kinds of main characteristics of nodes, we divide several basic centrality methods into four groups and design different combination schemes. Then, we use entropy weight method to weigh the attributes in each combination quantitatively. In contrast to other similar multi-attribute methods, our method emphasizes the differences of different attributes on nodes, and also avoids subjective weighting. Besides, the influence of neighbor nodes on the target nodes is abstractly expressed by the "gravity" between nodes. In other gravity centrality methods, only the nearest neighbor is considered, however, in our method the nearest, the second nearest and the third nearest neighbors are all considered. Last but not least, by comparing the results of different combinations, we choose the optimal combination as the basis of identifying influential nodes, thereby increasing the accuracy of ranking results.

3. Experiment

3.1. Data

We chose six real and typical networks to prove the validity and accuracy of our proposed method. The detailed descriptions for six networks are as follows: (i) Friendship:[37] This directed network represents the friendship between students. Nodes represent students and the edge weights indicate interactions. (ii) Advogato:[38] This is the trust network of Advogato. Nodes are the users of Advogato and the directed edges represent trust relationships. (iii) Trust:[39] This is a network of who-trusts-whom relationships among users of the Bitcoin Alpha platform. A weighted edge from i to j represents trust ratings. (iv) Collaboration:[40] This network is a collaboration network. Nodes are the authors and edges are joint work between authors. (v) Airlines:[41] This is the directed network of flights between US airports. Nodes represent airports and edges represent connections from one airport to another. (vi) WormNet:[42] This is a network by integration of all data-type-specific networks by modified Bayesian integration. Nodes represent the genes and edges represent lines. Several statistical features of six networks are listed in Table 2, including the number of nodes n, the number of edges m, average degree $\langle k\rangle$, largest degree $k_{\max}$, degree assortativity r, and the epidemic threshold $\beta_c \approx \langle k\rangle / \left(\langle k^2\rangle - \langle k\rangle\right)$.

Table 2. Basic statistical features of four networks.

Network	n	m	$\langle k\rangle$	$k_{\max}$	r	$\beta_{\rm c}$
Friendship	2539	12969	10.216	36	0.232	0.082
Advogato	5155	47135	18.287	941	−0.089	0.011
Trust	3783	24186	12.787	888	−0.163	0.010
Collaboration	7343	11898	3.864	102	0.243	0.070
Airlines	1574	28236	35.878	596	−0.122	0.005
WormNet	2220	53683	48.363	242	0.068	0.011

3.2. SIR model

In order to compare the ability to identify the influential nodes in complex networks by using different centrality methods, we use standard SIR[36] model to simulate the spreading process and evaluate the ability of different nodes to propagate. In the SIR model, each node has three states, *i.e.*, (i) susceptible state, $S(t)$ is used to represent the number of individuals that are susceptible to (not yet infected) the disease; (ii) infected state, $I(t)$ denotes the number of individuals which have been confirmed to be infected and can spread the disease to other susceptible individuals; (iii) recovered state, $R(t)$ denotes the number of previously infected individuals that have recovered already and will never be infected any more. In each experiment, only one node is infected at the initial time, and the others are the susceptible. At each time step, the infected node randomly infects its susceptible neighbors with the probability λ, and the infected node becomes the recovered with the probability β. When there are no more infected nodes in the network, the spreading process ends. At the moment t, the total number of infected and recovered nodes in the network is denoted as $F(t)$, which can be used to evaluate the transmission capacity of the initial infected node at time t. Obviously, in the ranking list, the more influential the node, the larger the value of $F(t)$ at time t will be. For each initial infected node, we carry on 300 simulations and take the average as the final simulation data.

3.3. Validation parameters

3.3.1. Complementary cumulative distribution function

In identifying influential nodes in complex networks, many nodes may have the same centrality values, *i.e.*, ranking values. For example, nodes in the same layer have the same k-shell value; nodes with the same neighbors have the same degree value. So it is difficult to distinguish the differences among these nodes. Here, we use the complementary cumulative distribution function (CCDF)[43] to analyze the distribution of influential nodes, and then make a comparison of advantage and disadvantage among various centrality methods. For good centrality methods, different nodes should have different ranking values, so the distribution of CCDF is relatively decentralized and the ranking range is larger than those of other methods.

$$\mathrm{CCDF}(z) = \mathrm{Prob}(Z > z) = 1 - \mathrm{CDF}(z), \tag{7}$$

where $\mathrm{CD}F(z)$ is the cumulative distribution function, also known as the distribution function, which is the integral of probability density function within the range of the variable Z less than specified value z. For discrete variables, it represents the sum of all values less than or equal to z ($\mathrm{CD}F(z) = \mathrm{Prob}(Z \leqslant z)$).

3.3.2. Kendall's tau coefficient

To assess the performances of different centrality methods, we use Kendall's tau coefficient[44–47] to analyze and verify the simulation results. The Kendall's tau coefficient is an index used to reflect the correlation between two ordered sequences. Here, a series of influential nodes is obtained based on the centrality values and another one is based on the SIR simulation results. The higher the value of Kendall's tau coefficient, the higher the correlation between the two ordered sequences is, that is, the more accurate the ranking list of the influential nodes generated by centrality methods is.

Definition 1 X and Y are two random variables, $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ are a set of observed values of random variables X and Y respectively. If $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$, (x_i, y_i) and (x_j, y_j) are condisered to be concordant. If $x_i > x_j$ and $y_i < y_j$, or $x_i < x_j$ and $y_i > y_j$, (x_i, y_i) and (x_j, y_j) are regarded as being discordant. If $x_i = x_j$ and

$y_i = y_j$, (x_i, y_i) and (x_j, y_j) are neither concordant nor discordant.

$$\tau = \frac{n_c - n_d}{0.5N(N-1)}, \tag{8}$$

where n_c and n_d represent the total number of the concordant and discordant observations,respectively, and N denotes the total number of nodes. The higher the value τ, the more accurate the ranking list of influential nodes is.

3.3.3. Imprecision function

The imprecision function is another method of assessing the accuracy of centrality methods, which is proposed by Kitsak *et al.*[7,48] and modified by Liu *et al.*[48,49] Based on the sequence of influential nodes generated by the centrality method and the sequence generated by the real SIR numerical simulation, the imprecision function is used to analyze the difference in average spreading capability between the top M nodes in two ordered sequences. The definition of this function is as follows:

$$\varepsilon(p) = 1 - \frac{S(p)}{S_{\text{eff}}(p)}, \tag{9}$$

where p represents the ratio of the number of top M nodes to the total number of nodes N $(p = M/N)$. $S(p)$ and $S_{\text{eff}}(p)$ represent the average spreading capability of the top M nodes in two ordered sequences determined by the centrality method and SIR numerical simulation respectively. The smaller the value ε, the closer to that determined by SIR numerical simulation the nodes sequence determined by the centrality method is, that is, the more accurate the ranking list of influential nodes is.

4. Results and analysis

4.1. Comparisons among multi-attribute combination methods

Based on the idea of the multi-attribute decision-making, in Section 2, we obtained nine combinations. Here we compare and analyze the differences among these combination methods in the identifying influential nodes, and select the optimal combination, namely our proposed centrality method.

4.1.1. Comparing correlation of different combinations

According to the definition of Kendall's tau coefficient, the larger the coefficient τ, the more accurate the ranking list of influential nodes is. We use Kendall's tau coefficient to analyze the correlations between node spreading capability $F(t)$ and centrality values of nodes in six real networks. The results are shown in Fig. 1. By changing the infection probability λ, we can obtain a series of Kendall's tau coefficients for each combination. First of all, we find that in each graph, the trends of all curves are basically the same, which shows that the nine combinations have certain similarity in identifying influential nodes. Secondly, the simulation results of combination 6 are relatively stable, and the curve is basically above all other curves as indicated in Figs. 1(a)–1(f), except for several points, for example, the value of λ is 0.15 in Fig. 1(a) and 0.03 in Fig. 1(e).

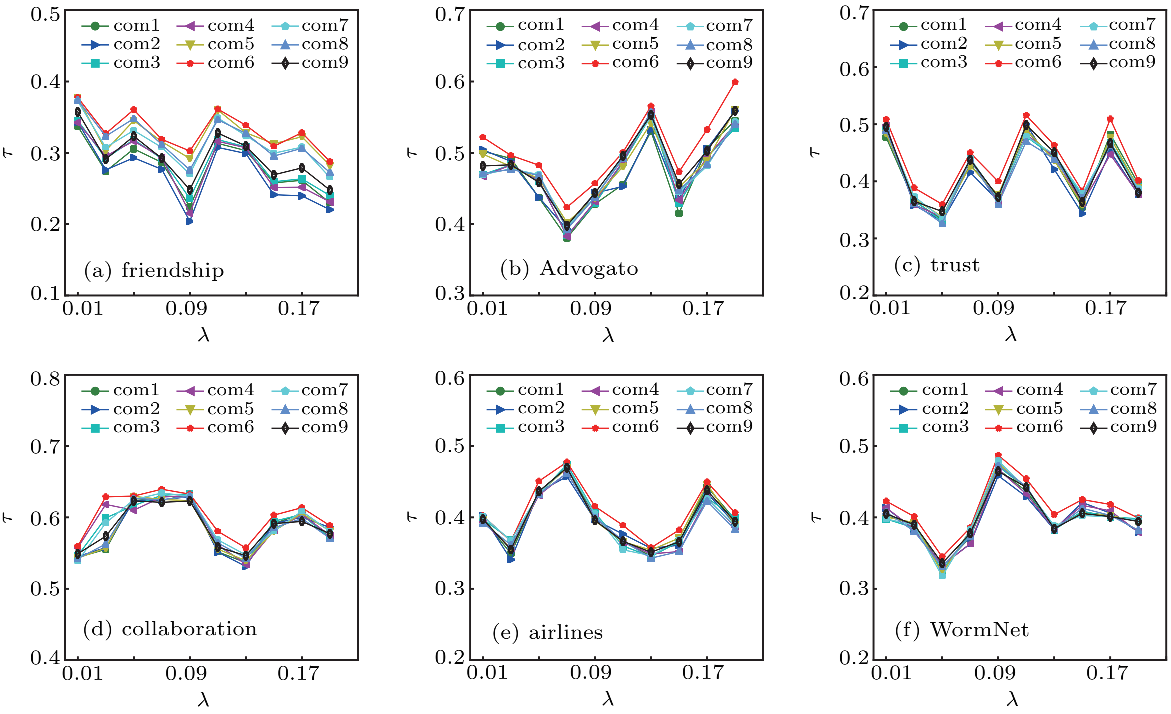

Fig. 1. Comparisons of ranking similarity between various combination methods and SIR model over six real-world networks. Kendall's tau coefficient τ is acquired by varying infection probability λ, and results are obtained by averaging over 300 independent realizations.

In addition, the value τ of combination 9 (com9) is not the largest in the six networks, that is, when all attributes are considered, the ranking result is not optimal. In general, combination 6 (com6) has the larger value of τ and more accurate ranking results in identifying influential nodes.

4.1.2. Comparisons of imprecision among different combinations

In the above, by comparing the coefficient τ, we know that the combination 6 is better than other combinations. Here, we choose the imprecise function to further prove this conclusion. The smaller the imprecise function ε, the more accurate the ranking result is. The results are shown in Fig. 2.

In Fig. 2, with the increase of the value p, all curves show a downward trend, showing that the larger the number of the top M nodes in ranking list, the smaller the value of ε is. In Figs. 2(a)–2(f), the curve of combination 6 is located below other curves and has a smaller value of ε, which shows that the sequence of influential nodes determined by combination 6 is closer to that determined by SIR simulation. In Fig. 2(e), when the value of p is 0.03, the value of ε is larger than that of p (0.01). It shows that the average accuracy of top 3% of all influential nodes is lower than that of top 1%. All in all, combination 6 has a smaller ε in six networks, which further proves that combination 6 is superior to other combinations.

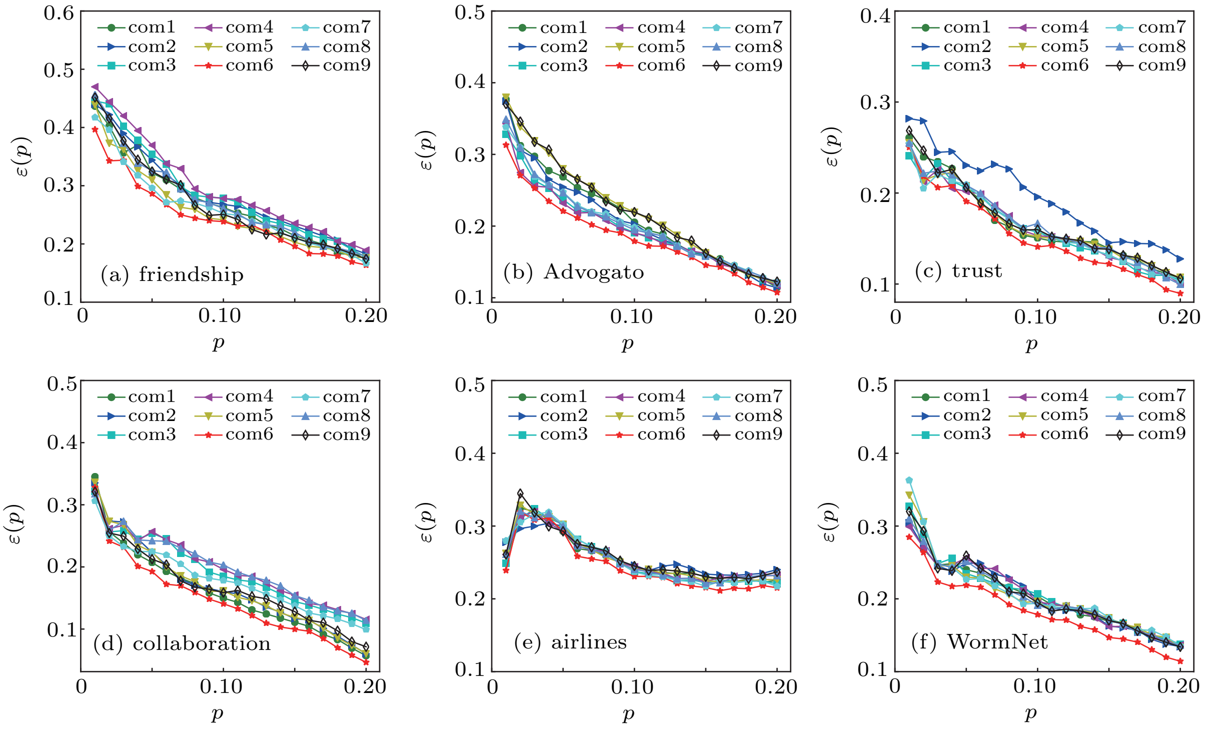

Fig. 2. Comparisons of ranking imprecision between various methods and SIR model over six real-world networks. Imprecision function ε is acquired by varying the fraction of network size p. Results are obtained by averaging over 300 independent realizations.

4.2. Comparisons between optimal combination method and seven centrality methods

Through the researches in Subsection 4.1, we find that among all combinations, combination 6, namely the multi-attribute combination of H-index, CC, BC, and PR, has more stable and accurate effect in identifying influential nodes. In this subsection, we mainly compare and analyze the differences between the combination 6 and seven basic centrality methods considered in the multi-attribute decision-making, to further determine the feasibility and accuracy of combination 6.

4.2.1. Comparisons of similarity ranking among all nodes

We use the comprehensive cumulative distribution function (CCDF) to compare the distribution of the ranking values for revealing the distinction of node between the optimal combination and seven basic centrality methods. The wider the range of ranking list, the better the centrality method is. The results are shown in Fig. 3. In Fig. 3, it can be clearly observed that the ranking ranges created by different centrality methods are different from each other. The distributions of the ranking values of EC, PR, and combination 6 are relatively wider than those of other centrality methods. By comparison, DC, k-shell, and H-index algorithm have small ranking ranges. Generally speaking, the ranking range created by combination 6 is much larger than by the seven basic centrality methods, and it shows our proposed method can more easily reveal a difference among the nodes within the network.

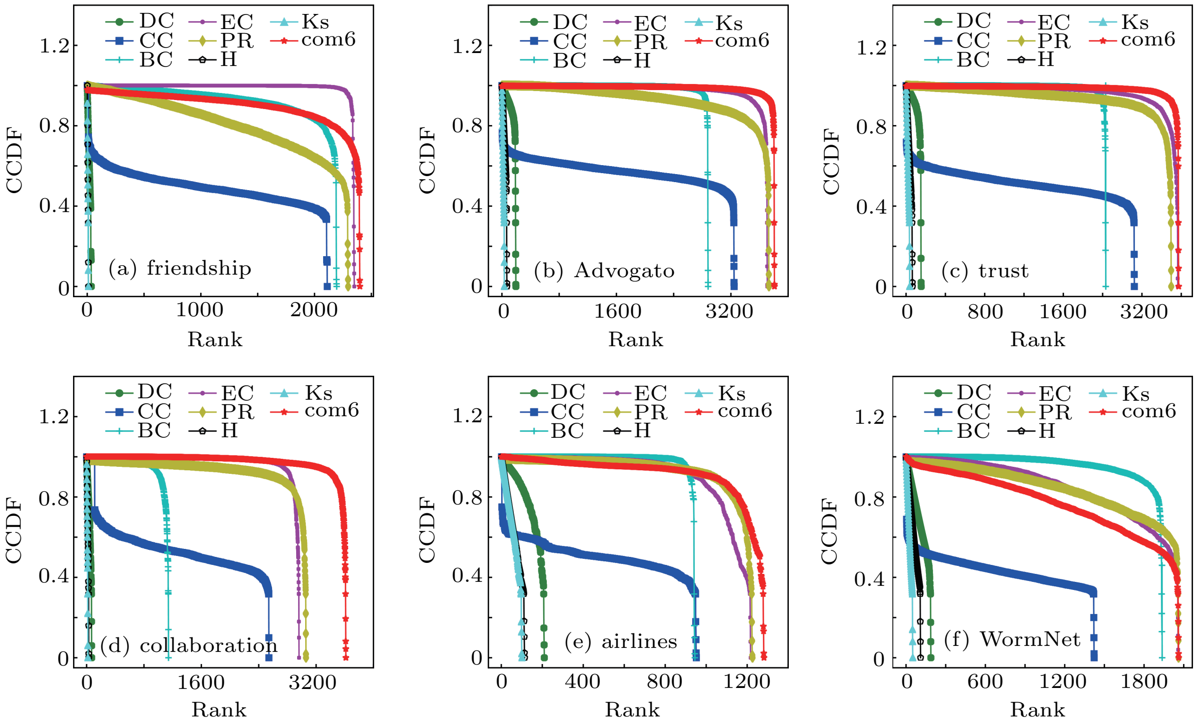

Fig. 3. Comparisons of complementary cumulative distribution function between optimal combination and seven basic centrality methods. Curves are acquired by sequencing all nodes based on CCDF, and results are obtained by averaging over 300 independent realizations.

4.2.2. Comparisons of correlation between optimal combination and centrality methods

A good centrality method should not only distinguish nodes as much as possible, but also rank influential nodes as accurately as possible. It is obvious that the curve generated by combination 6 is higher than those generated by other centrality methods especially in Figs. 4(b) and 4(d). It means that our method performs much better than the other centralities in all six networks. In Fig. 4(f), all curves have consistent trends and similar values of τ, which indicates the sequences of influential nodes generated by these centrality methods are relatively uniform in the network of WormNet. We can also find that these traditional centrality methods do not perform consistently in six networks. For example, the BC in Fig. 4(a) has larger values of τ than most of methods, but smaller values of τ in Figs. 4(d), 4(e), and 4(f).

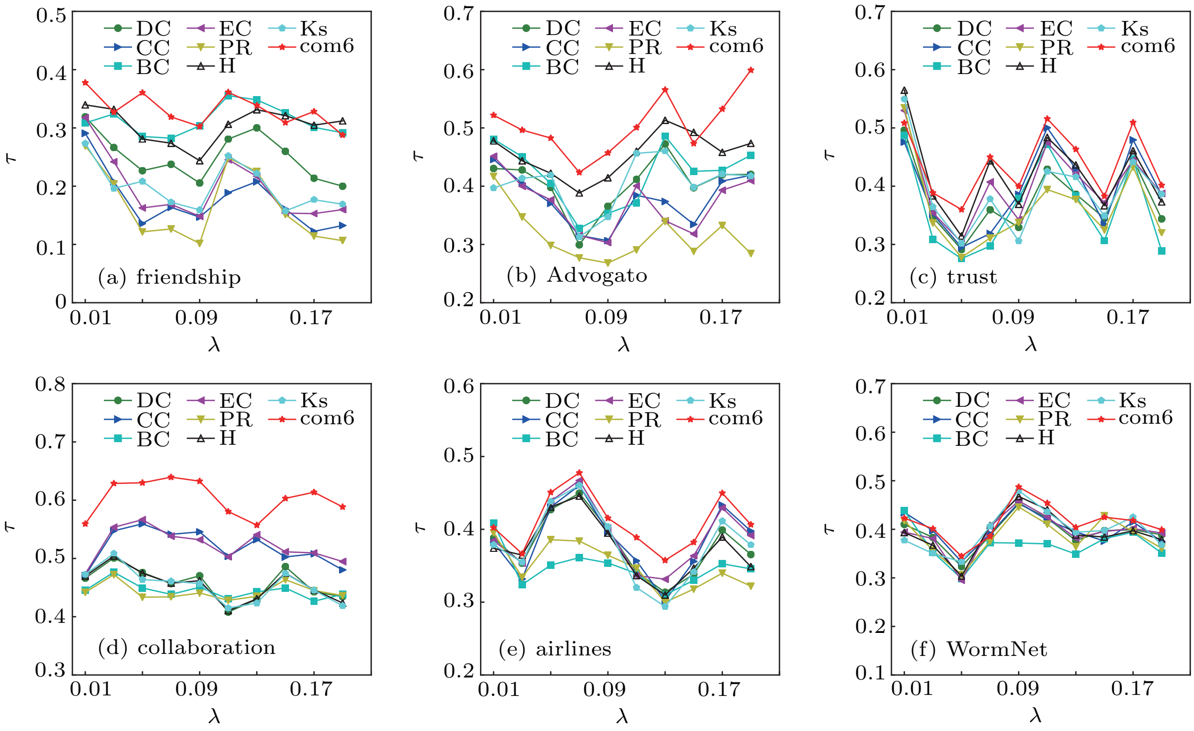

Fig. 4. Comparisons of Kendall's tau coefficient τ between optimal combination (red curve) and seven centrality methods according to SIR model in six real networks. Each point is obtained by averaging over 300 independent runs.

Thus these traditional centrality methods are hard to apply to the most networks with different structures. Therefore, our method has more stable and accurate performances.

4.2.3. Comparisons of imprecision between optimal combination and centrality methods

According to the definition of the imprecision function, the smaller the value of ε, the more accurate the centrality method is. In Fig. 5, with the increase of p, the imprecision of most of centrality methods has a consistent downward trend. In Fig. 5(d), that is, in the network of collaboration, the imprecisions of the EC, CC, and combination 6 are basically the same, but are significantly smaller than those from other centrality methods. All in all, the curve created by combination 6 has lower imprecision, except for at a few points. It means that our method can not only identify influential nodes accurately, but also be widely applied to different complex networks. Therefore, our novel method is superior to the previous centrality methods, which is of great significance for operating and controlling the complex networks.

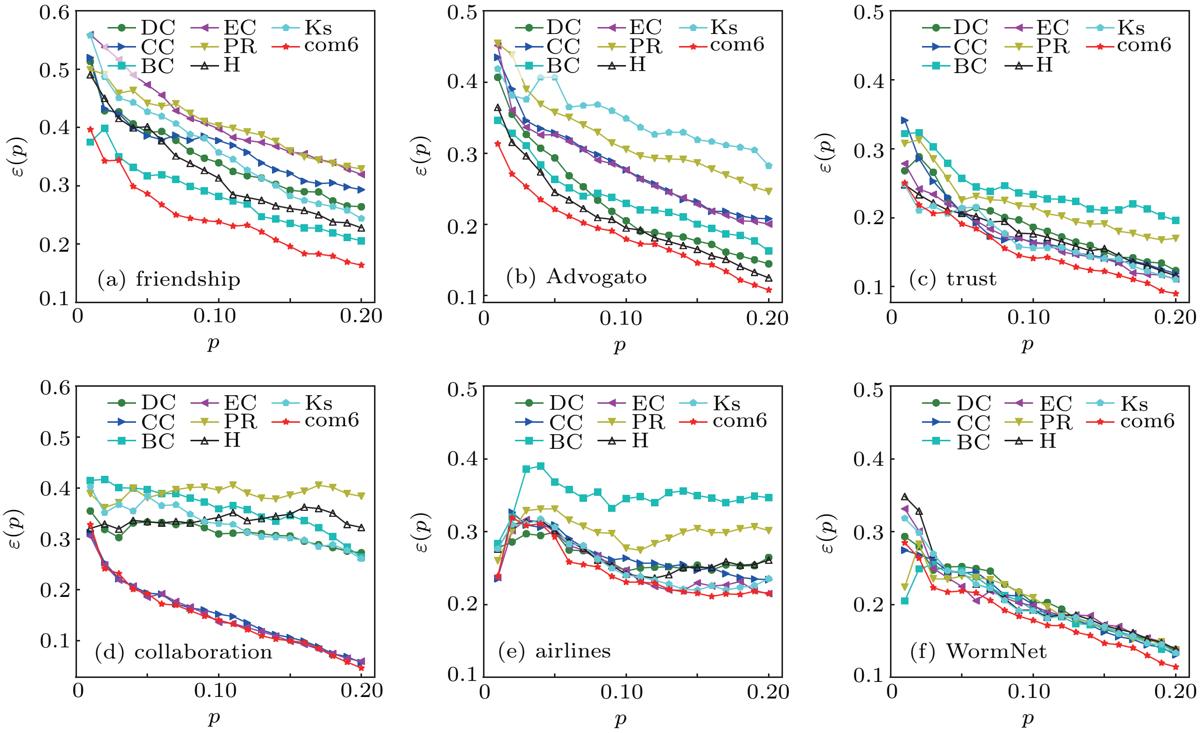

Fig. 5. Comparisons of imprecision between the optimal combination (red curve) and seven basic centrality methods over six real-world networks. Results are obtained by averaging over 300 independent realizations.

5. Conclusions and discussion

In this paper, we proposed a novel centrality method of identifying influential nodes in the complex networks. Considering the shortcomings of the traditional centrality methods, we adopt the idea of multi-attribute decision-making to comprehensively consider the multiple characteristics of nodes, and use entropy weight method to quantitatively weigh different attributes to obtain the combined centrality of nodes. Then, according to the gravity law, we replace the mass with the combined centrality to calculate the ‘gravity’ of the target node, and take this ‘gravity’ as the basis for identifying influential nodes. In order to evaluate the performances, we apply our method to six real networks and use SIR model to simulate the spreading process. By calculating the comprehensive cumulative distribution function (CCDF), we find that the proposed method can better distinguish nodes. Then, by calculating Kendall’s tau coefficient τ and the imprecision function ε, we find that ranking results of influential nodes generated by proposed method are more stable and accurate in different complex networks.

After that, further efforts are needed to improve the proposed method. Firstly, the selection of the basic attributes of nodes is relatively subjective. So how to rationally select attributes according to the real situation of networks is very necessary. Then, we use the classical SIR model to simulate spreading process. However, in the real-society networks with complex structures, nodes do not completely evolve according to the SIR model. Therefore, a more accurate and appropriate numerical model may be one of the important challenges and breakthroughs in the research of identifying influential nodes in complex networks.

References

[1] Zanin M and Lillo F 2013 *Eur. Phys. J. Spec. Top.* **215** 5

[2] Lordan O, Sallan J M and Simo P 2014 *J. Transp. Geogr.* **37** 112

[3] Li H J, Li H Y and Jia C L 2015 *Int. J. Mod. Phys. C* **26** 1550043

[4] Arularasan A N, Suresh A and Seerangan K 2019 *Cluster Comput.* **22** 4035
[5] Shang Y L 2015 *J. Syst. Sci. Complexity* **28** 96
[6] Ma L L, Ma C, Zhang H F and Wang B H 2016 *Physica A* **451** 205
[7] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E and Makse H A 2010 *Nat. Phys.* **6** 888
[8] Bae J and Kim S 2014 *Phys. A Stat. Mech. Appl.* **395** 549
[9] Lü L Y, Zhang Y C, Yeung C H and Zhou T 2011 *PLoS ONE* **6** e21202
[10] Chen D B, Lü L Y, Shang M S, Zhang Y C and Zhou T 2012 *Physica A* **391** 1777
[11] Lü L Y, Chen D B and Zhou T 2011 *New J. Phys.* **13** 123005
[12] Mehta A and Gupta R 2015 arXiv:1509.07966v1 [cs.SI]
[13] Wang X Y, Wang Y, Qin X M, Li R and Eustace J 2018 *Chin. Phys. B* **27** 100504
[14] Fei L G and Deng Y 2017 *Chaos, Solitons and Fractals* **104** 257
[15] Kang L, Xiang B B, Zhai S L, Bao Z K and Zhang H F 2018 *Acta Phys. Sin.* **67** 198901 (in Chinese)
[16] Freeman L C 1978 *Soc. Netw.* **1** 215
[17] Freeman L C 1977 *Sociometry* **40** 35
[18] Bonacich P and Lloyd P 2001 *Soc. Netw.* **23** 191
[19] Lü L Y, Chen D B, Ren X L, Zhang Q M, Zhang Y C and Zhou T 2016 *Phys. Rep.* **650** 1
[20] Wen T and Deng Y 2020 *Inform. Sci.* **512** 549
[21] Fei L G, Zhang Q and Deng Y 2018 *Physica A* **512** 1044
[22] Gao S, Ma J, Chen Z M, Wang G H and Xing C M 2014 *Physica A* **403** 130
[23] Zhong L F, Liu J G and Shang M S 2015 *Phys. Lett. A* **379** 2272
[24] Wang Y C, Wang S S and Deng Y 2019 *Pramana - J. Phys.* **92** 68
[25] Zeng A and Zhang C J 2013 *Phys. Lett. A* **377** 1031
[26] Song B, Jiang G P, Song Y R and Xia L L 2015 *Chin. Phys. B* **24** 100101
[27] Yin R R, Yin X L, Cui M D and Xu Y H 2019 *J. Wireless Com. Network* **2019** 234
[28] Fei L G, Mo H M and Deng Y 2017 *Mod. Phys. Lett. B* **31** 1750243
[29] Du Y X, Gao C, Hu Y, Mahadevan S and Deng Y 2014 *Physica A* **399** 57
[30] Liu Y J, Wu J and Liang C Y 2015 *Kybernetes* **44** 1437
[31] Mo H M and Deng Y 2019 *Physica A* **529** 121538
[32] Bian T, Hu J T and Deng Y 2017 *Physica A* **479** 422
[33] Hu J T, Du Y X, Mo H M, Wei D J and Deng Yong 2016 *Physica A* **444** 73
[34] Li Z, Ren T, Ma X Q, Liu S M, Zhang Y X and Zhou T 2019 *Sci. Rep.* **9** 8387
[35] Ibnoulouafi A and Haziti M E 2018 *Chaos, Solitons and Fractals* **114** 69
[36] Kermack W O and McKendrick A G 1927 *Proc. R. Soc. Lond. A* **115** 700
[37] http://konect.uni-koblenz.de/networks/moreno_health
[38] http://konect.uni-koblenz.de/networks/advogato
[39] https://icon.colorado.edu/#!/networks
[40] http://vlado.fmf.uni-lj.si/pub/networks/data/collab/geom.htm
[41] http://konect.uni-koblenz.de/networks/opsahl-usairport
[42] http://networkrepository.com/bio-CE-GN.php
[43] Li C, Wang L, Sun S W and Xia C Y 2018 *Appl. Math. Comput.* **320** 512
[44] Knight W R 1966 *J. Amer. Statist. Assoc.* **61** 436
[45] Kendall M G 1938 *Biometrika* **30** 81
[46] Kendall M G 1945 *Biometrika* **33** 239
[47] Ruan Y R, Lao S Y, Xiao Y D, Wang J D and Bai L 2016 *Chin. Phys. Lett.* **33** 028901
[48] Wang J Y, Hou X N, Li K Z and Ding Y 2017 *Physica A* **475** 88
[49] Liu Y, Tang M, Zhou T and Do Y H 2015 *Sci. Rep.* **5** 9602