Identifying Key Nodes in Complex Networks Based on Global Structure

YUANZHI YANG1, XING WANG1, YOU CHEN1, AND MIN HU2
1Aeronautics Engineering College, Air Force Engineering University, Xi’an 710038, China
2School of Materials Science and Engineering, Southwest Petroleum University, Chengdu 610500, China

Corresponding authors: You Chen (chenyousky@163.com) and Min Hu (hu_min_min@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 61472443.

ABSTRACT
Quantitative identification of key nodes in complex networks is of great significance for studying the robustness and vulnerability of complex networks. Although various centralities have been proposed to solve this issue, each approach has its limitations for its own perspective of determining an actor to be “key”. In this paper, we propose a novel method to identify key nodes in complex networks based on global structure. Three aspects including the shortest path length, the number of shortest paths and the number of non-shortest paths are considered, and we establish three corresponding influence matrices. Node efficiency, which can reflect the contribution of one node to the information transmission of the entire network, is selected as the initial value of node’s influence on other nodes, and then the comprehensive influence matrix is constructed to reflect the influence among nodes. The proposed method provides a new measure to identify key nodes in complex networks from the perspective of global network structure, and can obtain more accurate identification results. Four experiments are conducted to evaluate the performance of our proposed method based on Susceptible–Infected (SI) model, and the results demonstrate the superiority of our method.

INDEX TERMS
Complex networks, key nodes, global structure, SI model.

I. INTRODUCTION
Complex network is the abstract expression of a real complex system, in which elements are abstracted as nodes, and the relationships between elements are abstracted as edges between nodes. Complex networks have heterogeneous topology, which makes it impossible for each node in network to have the same importance [1]. Therefore, it is of great theoretical significance and application prospect to identify key nodes in complex networks by quantitative methods and utilize the properties of key nodes [2]. For example, stopping the spread of rumors and viruses [3], [4], drug targeting and discovery in biomedicine [5], [6], guiding effective information to spread rapidly in the network [7], [8], deliberate attack on terrorist and drug networks [9].

With the development of complex networks, a variety of methods have been put forward from different perspectives, such as degree centrality (DC) [10], closeness centrality (CC) [11], betweenness centrality (BC) [12] and Katz centrality [13], etc. Degree centrality only considers the information of the target node and its neighbors, and it has low accuracy and low time complexity. Closeness centrality and betweenness centrality are well-known global centralities which have high accuracy and time complexity, thus the two centralities cannot be applied to large-scale networks. To achieve a balance between high accuracy and low time complexity, Chen et al. [14] came up with a semi-local centrality (LC) based on multi-level neighbor information to rank nodes. This method considers not only the degree of nodes but also the neighborhood information, but does not consider the topology connections among the neighbors. The centralities mentioned above ignore the effect of location of nodes in the network, Kitsak et al. [15] believed that the influence of nodes was related to the location of nodes in the network, they proposed to use K-shell (Ks) decomposition to identify key nodes from a new perspective. This method considers that the importance of key nodes is related to the location of the network. The outer nodes are stripped layer by layer, and the nodes in the inner layer are the key nodes in the network. But this method will result in a large number of nodes being assigned the same Ks value. Then a series of methods to expand and improve K-shell decomposition were put forward. Bae and Kim [16] considered that the sum of the Ks value of nodes and their neighbors should be used to measure the importance...
of nodes. Liu et al. [17] comprehensively considered the Ks value of target node and the distance between it and the node with maximum Ks value in the network. Zeng and Zhang [18] proposed a mixed degree decomposition method, which considered the degree information of removing nodes in calculation, and obtained high ranking accuracy of influence. In the field of search engine, there are also some methods such as PageRank [19], LeaderRank [20] and Hits [21]. In addition, information entropy has been utilized to rank the influence of nodes in complex networks [22]–[25]. Some researchers [26]–[29] proposed to combine indicators with multi-attribute ranking methods.

In this paper, we turn a sight to the global network structure, which considers the shortest path length, the number of shortest paths and the number of non-shortest paths. And three corresponding influence matrices (efficiency matrix, direct arrival paths matrix and indirect arrival paths matrix) are constructed and considered to form the comprehensive influence matrix. Besides, node efficiency is selected as the initial value of node’s influence on other nodes. We can identify key nodes effectively by considering these three aspects comprehensively. In order to evaluate the feasibility and effectiveness of the proposed method, Susceptible–Infected (SI) model [30] is utilized to simulate the spreading process in network, and four experiments are conducted based on four real datasets to verify the effectiveness of our method. Other classical centralities (DC, CC, BC, Ks and LC) are employed to be compared with our method in four aspects, and the experimental results demonstrate the superiority of the proposed method.

The remaining of this paper is organized as follows. In section 2, we introduce some related work and classical centralities. Our method is proposed in section 3. In section 4, we conduct four experiments to evaluate the performance of the proposed method based on four real datasets. Finally, section 5 gives a conclusion.

II. RELATED WORK

Suppose that an undirected and unweighted network $G(V, E)$ consists of $m = |E|$ edges and $n = |V|$ nodes, and the adjacent matrix $A = (a_{ij})_{n \times n}$ is used to describe the network, which is defined as

$$a_{ij} = \begin{cases} 1, & \text{node } i \text{ is connected with } j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Degree centrality [10] is the simplest centrality in identifying key nodes. Degree centrality of one node is defined as the number of its nearest neighbors. A high value of degree centrality indicates that one node is able to affect numerous neighbors directly. It is denoted as

$$DC(i) = \sum_{j=1}^{n} a_{ij} \quad (2)$$

Closeness centrality [11] uses the shortest paths between all pairs of nodes to determine the influence, it is a global centrality with high time complexity. It is defined as the derivative of the average shortest distance from a node to others in the network.

$$CC(i) = \frac{n - 1}{\sum_{j=1}^{n} d_{ij}} \quad (3)$$

wherein, $d_{ij}$ represents the distance between node $i$ and node $j$.

Betweenness centrality [12] believes that a node is an influential one if the number of shortest paths passing through the node is great. It describes the influence of a node on information flow in network.

$$BC(i) = \sum_{s \neq i \neq t} \frac{g_{st}(i)}{g_{st}} \quad (4)$$

wherein, $g_{st}$ denotes the number of all the shortest paths between node $s$ and node $t$, $g_{st}(i)$ represents the number of shortest paths passing through node $i$.

K-shell decomposition [15] considers the location of nodes in the network. In this approach, the outer nodes are stripped layer by layer, and the inner nodes have high influence. The decomposition process assigns a $k\text{s}$ value to each node. It can be regarded as a coarse-grained ranking method based on node degree.

Local centrality [14] achieves a trade-off between accuracy and time complexity. It takes the nearest and the next nearest neighbors into consideration.

$$Q(j) = \sum_{w \in \Gamma_j} N(w) \quad (5)$$

$$C_L(i) = \sum_{j \in \Gamma_i} Q(j) \quad (6)$$

wherein, $\Gamma_i$ represents the set of the nearest neighbors of node $i$ and $N(w)$ denotes the total number of the nearest and the next nearest neighbors of node $w$.

Node efficiency is the average value of the sum of the reciprocal distances between the node and other nodes in the network, it can represent the ability from one node to other nodes in the network. It is defined as

$$I(i) = \frac{1}{n - 1} \sum_{j \neq i, j=1}^{n} \frac{1}{d_{ij}} \quad (7)$$
Y. Yang et al.: Identifying Key Nodes in Complex Networks Based on Global Structure

FIGURE 2. The frequency of nodes with the same ranking value in four networks. (a) In Karate club network. (b) In Jazz musicians network. (c) In USAir97 network. (d) In Email network.

III. THE PROPOSED METHOD

We turn a sight to the global structure of network, which can provide a more accurate approach for key nodes identification. Nodes in a network are not isolated, but affected and restricted by other nodes. The relationship among nodes can be described by the influence matrix. From the perspective of information transmission path, the influence among nodes will be affected by two factors: the shortest path length and the number of shortest paths. It should be noted that node can also transmit its importance through non-shortest paths, thus affecting the importance of the pointed node. Therefore, we comprehensively consider these three aspects (the shortest path length, the number of shortest paths and the number of non-shortest paths) and construct three corresponding influence matrices.

A. THE INFLUENCE MATRIX BASED ON EFFICIENCY

According to the theory of spatial autocorrelation, it can be considered that the influence between two nodes is inversely proportional to the distance between them. The efficiency between two nodes can be defined as

$$e_{ij} = \frac{1}{d_{ij}}$$  \hspace{1cm} (8)

When there is no path between node pairs, we define $d_{ij} = +\infty$, which will make it impossible to use $d_{ij}$ representation directly. Hence, we use $e_{ij}$ (derivative of $d_{ij}$) to avoid this issue. If there is no path between node $i$ and $j$, $e_{ij} = 0$. If the two nodes are connected directly, $e_{ij} = 1$. Then the influence matrix based on efficiency (abbreviated as IME) can be established as follows.

$$IME = \begin{bmatrix}
0 & e_{12} & \cdots & e_{1n} \\
e_{21} & 0 & \cdots & e_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
e_{n1} & e_{n2} & \cdots & 0
\end{bmatrix}$$  \hspace{1cm} (9)

IME can reflect the influence between nodes from the perspective of the shortest path length.
FIGURE 3. The average spreading ability of top-10 nodes ranked by different methods. The maximum time step is set as \( t = 50 \) and the simulation runs 1000 Monte Carlo. (a) In Karate club network. (b) In Jazz musicians network. (c) In USAir97 network. (d) In Email network.

B. THE INFLUENCE MATRIX BASED ON THE NUMBER OF SHORTEST PATHS

In this part, we establish the influence matrix based on the number of shortest paths. The influence between two nodes is also affected by the number of shortest paths. Take FIGURE 1 as a simple example, we focus on the influence of node 6 affected by node 3 and node 10. The number of shortest path between node 3 and node 6 is 2, which is the same as that between node 10 and node 6. The number of shortest paths between node 3 and node 6 is 1 (3-5-6), but the number of shortest paths between node 10 and node 6 is 3 (10-7-6, 10-8-6 and 10-9-6). Obviously, the influence of node 10 on node 6 is greater than that of node 3 on node 6.

The number of shortest paths between two nodes can be calculated by the adjacent matrix \( A \). For any positive integer \( k \) \((k > 2)\), the element value \((A^k)_{ij}\) in the \( k \)-th power of \( A \) can be calculated as

\[
(A^k)_{ij} = \sum_{m_1=1}^{n} \sum_{m_2=1}^{n} \cdots \sum_{m_{k-2}=1}^{n} \sum_{m_{k-1}=1}^{n} a_{im_1}a_{m_1m_2} \times \cdots \times a_{m_{k-2}m_{k-1}}a_{m_{k-1}j} \quad (10)
\]

The element \((A^k)_{ij}\) represents the total number of paths with length \( k \) between node \( i \) and node \( j \). Therefore, we can obtain the number of shortest paths between node \( i \) and node \( j \), that is to say, the number of paths with length \( d_{ij} \) between node \( i \) and node \( j \) is \((A^d)_{ij}\). Then the influence matrix based on the number of shortest paths (abbreviated as IMS) can be established as (11), shown at the bottom of the next page, IMS can reflect the influence between nodes from the perspective of the number of shortest paths. Essentially, the element in IMS represents the ability of nodes to transmit information along the shortest path.
FIGURE 4. The Kendall’s tau coefficient between six methods and the actual spreading situation $\varepsilon$. The ranking list $\varepsilon$ is simulated by SI model at $t = 20$. (a) In Karate club network. (b) In Jazz musicians network. (c) In USAir97 network. (d) In Email network.

C. THE INFLUENCE MATRIX BASED ON THE NUMBER OF NON-SHORTEST PATHS

Sometimes, information is not transmitted along the shortest path, and the influence of the number of non-shortest paths also needs to be considered. For example, we focus on the influence of node 5 affected by node 3 and node 6 in figure 1. The shortest path length, as well as the number of shortest paths, between node 3 and node 5 is the same as that between node 6 and node 5. Next, we consider the number of non-shortest paths. When the path length is 2, there is no path between node 3 and node 5, or between node 6 and node 5. When the path length is 3, the number of

$$IMS = \begin{bmatrix}
0 & A_{d_{12}}^{12} + A_{d_{12}}^{22} + \cdots + A_{d_{12}}^{n_2} & \cdots & A_{d_{1n}}^{1n} + A_{d_{1n}}^{2n} + \cdots + A_{d_{1n}}^{nn} \\
A_{d_{21}}^{d_{21}} + A_{d_{21}}^{d_{21}} + \cdots + A_{d_{n1}}^{d_{n1}} & 0 & \cdots & A_{d_{n2}}^{d_{n2}} + A_{d_{n2}}^{d_{n2}} + \cdots + A_{d_{n2}}^{nn} \\
\vdots & \vdots & \ddots & \vdots \\
A_{d_{n1}}^{d_{n1}} + A_{d_{n1}}^{d_{n1}} + \cdots + A_{d_{n1}}^{d_{n1}} & A_{d_{n2}}^{d_{n2}} + A_{d_{n2}}^{d_{n2}} + \cdots + A_{d_{n2}}^{nn} & \cdots & 0
\end{bmatrix} \quad (11)$$
paths between node 3 and node 5 is the same as that between node 6 and node 5. When the path length is 4, no path exists between node 3 and node 5, but there are four paths between node 6 and node 5 (6-7-8-6-5, 6-9-8-6-5, 6-8-7-6-5 and 6-8-9-6-5). We can find that the influence of node 6 on node 5 is greater than the influence of node 3 on node 5. Therefore, we consider the influence of the total number of non-shortest paths, which is defined as

\[ f_{ij} = \sum_{l=d_{ij}+1}^{d_{ave}} A^l_{ij} \] (12)

wherein, \( d_{ave} \) is the average path length of network. The path length between two nodes cannot increase indefinitely, and for the sake of calculation, we set the maximum path length as \( d_{ave} \). Then the influence matrix based on the number of non-shortest paths (abbreviated as IMN) can be established as (13), shown at the bottom of this page.

IMN can reflect the influence between nodes from the perspective of the number of non-shortest paths. Essentially, the element in IMN represents the ability of nodes to transmit information along the non-shortest path.

D. THE COMPREHENSIVE INFLUENCE MATRIX

From the above analysis, we can see that the influence between nodes is affected by the shortest path length, the number of shortest paths and the number of non-shortest paths. Therefore, we consider these three aspects and obtain an influence matrix as follows.

\[ IM = IME + IMS + IMN \] (14)

The element \( IM_{ij} \) in IM represents the influence of node \( i \) on node \( j \). Since node efficiency can reflect the contribution of one node to the information transmission of the entire network, node efficiency is selected as the initial value of node’s influence on other nodes. And we establish the comprehensive influence matrix.

\[ CIM = \begin{bmatrix} 0 & I_1M_{12} & \cdots & I_1M_{1n} \\ I_2M_{21} & 0 & \cdots & I_2M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ I_nM_{n1} & I_nM_{n2} & \cdots & 0 \end{bmatrix} \] (15)

Then the influence of node \( j \) can be expressed as

\[ D_j = \sum_{i=1}^{n} CIM_{ij} \] (16)

And the normalized influence of node \( j \) is denoted as

\[ N_j = \frac{D_j}{\sum_{k=1}^{n} D_k} \] (17)

IV. EXPERIMENTAL ANALYSIS

A. DATASETS

To evaluate the performance of the proposed method, we choose four datasets with varying sizes as the basis for experimental simulation and analysis. (i) Karate club [34]: It is a well-known social network with 34 members in a karate club in a US university, available online at “http://www-personal.umich.edu/~mejn/netdata/”. (ii) Jazz musicians [35]: Jazz musicians network is also a social network among jazz musicians, each node represents a jazz musician, and each edge indicates that two musicians have cooperation. It can be downloaded from “http://konect.uni-koblenz.de/networks/”. (iii) USAir97 [36]: It’s a transportation network consisting of 332 airports and 2126 routes. It is available online at “http://vlado.fmf.uni-lj.si/pub/networks/pajek/data/gphs.htm”. (iv) Email [37]: It is a well-known email network of the University at Rovira i Virgili, each node in the network represents a user and an edge indicates that two users have email exchange. The data can be obtained on “http://deim.urv.cat/~alexandre.arenas/data/welcome.htm”.

The basic statistics of four real datasets are presented in TABLE 1.

| Network        | \( N \) | \( E \) | \(<K>\) | \( C \) | \(<d>\) |
|----------------|--------|--------|--------|--------|--------|
| Karate club    | 34     | 78     | 4.59   | 0.59   | 2.41   |
| Jazz musicians | 198    | 2742   | 27.70  | 0.63   | 2.21   |
| USAir97       | 332    | 2126   | 12.81  | 0.75   | 2.74   |
| Email         | 1133   | 5451   | 9.62   | 0.22   | 3.61   |

B. SI MODEL

We adopt SI model [30] to simulate the spreading process in network. There are two states of nodes in SI model, including susceptible state and infected state. Nodes in susceptible state will be infected by nodes in infected state with a certain probability. Once a node is infected, it will never be recovered. Suppose that node \( i \) is the source node and the number of infected nodes will reach \( n_{it} \) after \( t \) (\( t = 1, 2, \ldots \)) time step.

\[ IMN = \begin{bmatrix} 0 & f_{i2} & \cdots & f_{in} \\ f_{21} & f_{12} + f_{22} + \cdots + f_{n2} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{1n} & \cdots & f_{nn} \end{bmatrix} \] (13)
Then the spreading ability of node $i$, denoted as $I_i(t) = n_i(t)/n$, can be defined as ratio of infected nodes to the total number of nodes. The maximum time step is set as $t = 50$ and the spreading process is conducted with 1000 Monte Carlo simulations. Obviously, the higher value of $I_i(t)$ indicates greater influence.

### C. THE KENDALL’S TAU COEFFICIENT

The Kendall’s tau coefficient [38] is a non-parametric statistic used to measure the degree of correspondence between two rankings and assessing the significance of this correspondence. Suppose $X$ and $Y$ are the ranking lists of two centralities, a pair of distinct nodes $i$ and $j$ is concordant if $((X_i > X_j)$ and $(Y_i > Y_j))$ or $((X_i < X_j)$ and $(Y_i < Y_j))$. If $((X_i > X_j)$ and $(Y_i < Y_j))$ or $((X_i < X_j)$ and $(Y_i > Y_j))$, the pair is discordant. The Kendall’s tau coefficient $\tau$ is defined as

$$\tau = \frac{n_c - n_d}{0.5 \times n \times (n - 1)}$$  \hspace{1cm} (18)

wherein, $n_c$ and $n_d$ are the number of concordant pairs and discordant pairs, respectively. Better performance of centrality has higher $\tau$ value.

### D. EFFECTIVENESS

We compare the performance of the proposed method with classical centralities including DC, CC, BC, Ks and LC. SI model [30] is adopted to simulate the spreading process in network with a certain spreading probability. In addition, the Kendall’s tau coefficient is utilized to evaluate the performance of different methods.

1) EXPERIMENT 1: COMPARE THE TOP-10 NODES USING DIFFERENT METHODS

In this experiment, we focus on the capability of identifying a group of key nodes. The top-10 ranked nodes are considered here. We rank the influence of nodes in four networks using classical five centralities and our proposed method, and the actual spreading ability $I = (I_1, I_2, \ldots, I_n)(t = 25)$ ranked by SI model is adopted for comparison. TABLE 2-5 show the top-10 nodes in four networks ranked by different methods and SI model.

According to TABLE 2-5, in Karate club network, the top-10 lists of the proposed method and CC are exactly the same as those of $I$. DC, BC and LC have 9 same nodes as $I$ in the top-10 lists, and Ks has 8 same nodes as $I$. In Jazz musicians network, the proposed method has 8 same nodes as $I$ in the top-10 lists, which is slightly worse than that of CC. There are 7 same nodes in the top-10 lists between DC, LC and $I$, BC has 5 same nodes, while Ks has the least number of same nodes. In USAir97 network, the proposed method and DC have 9 same nodes as $I$ in the top-10 lists, which are slightly greater than BC, CC and LC. Ks only has 3 same nodes as $I$ in the top-10 lists. As for Email network, the number of nodes in the top-10 lists of DC, CC, BC, Ks, LC and the proposed method that are the same as that of $I$ is 7, 8, 7, 0, 6, 8, respectively. In a word, the proposed method and CC have the largest number of nodes in the top-10 lists that are same as $I$, but Ks has the least number of nodes that are the same as $I$.

2) EXPERIMENT 2: COMPARE THE DISTINGUISHING CAPABILITY OF DIFFERENT METHODS

In experiment 1, we find a phenomenon that some nodes are assigned same ranking value, then these nodes cannot be
TABLE 6. The DIS value of different methods in four networks.

| Network       | DC     | CC     | BC     | Ks     | LC     | Proposed |
|---------------|--------|--------|--------|--------|--------|----------|
| Karate club   | 0.8235 | 0.5882 | 0.4706 | 0.9706 | 0.3235 | 0.3235   |
| Jazz musicians| 0.9343 | 0.5706 | 0.1465 | 0.9949 | 0.0707 | 0.5005   |
| USAir97       | 0.9217 | 0.6535 | 0.4970 | 0.9640 | 0.2620 | 0.2530   |
| Email         | 0.9938 | 0.4572 | 0.2189 | 1.0000 | 0.0688 | 0.0424   |

distinguished. Especially for Ks, it has the largest number of nodes with the same ranking value in the four networks. The frequency (the ratio of the number of nodes with the same ranking value to the number of total nodes) can be taken as an indicator to measure the distinguishing capability of a method, lower frequency of one method indicates better distinguishing capability. The frequency comparison results are presented in FIGURE 2. According to FIGURE 2, almost in all cases, the frequency of DC and Ks is much higher than other methods, that is to say, DC and Ks have the worst capability of distinguishing nodes. On the contrary, the proposed method has the lowest frequency almost in the four datasets and can further distinguish nodes. In addition, we put forward a parameter to further compare the distinguishing capability of different methods. It is defined as

\[
\text{DIS} = \frac{n - n_s}{n}
\]

wherein, \(n_s\) is the number of nodes with single ranking value. The smaller the DIS value, the better the distinguishing capability of a method. As shown in TABLE 6, we can find that the DIS value of the proposed method is the lowest almost in four networks except that LC and the proposed method both have the smallest DIS value in Karate club network, and the DIS value of DC and Ks is much greater than that of others. The proposed method has the best distinguishing capability.

3) EXPERIMENT 3: COMPARE THE AVERAGE SPREADING ABILITY OF TOP-10 NODES

In this part, we focus on the spreading ability of a group of nodes. A node is selected from the top-10 lists and taken as a source node. The average spreading ability of top-10 nodes can be defined as

\[
\text{ABILITY OF TOP-10 NODES} = \frac{\sum_{i=1}^{10} I(t(i))}{10}
\]

The maximum time step is set as \(t = 50\) and the simulation runs 1000 Monte Carlo. The average spreading ability of top-10 nodes ranked by different methods is presented in FIGURE 3.

As shown in FIGURE 3, In Karate club network, the curve of the proposed method overlaps the curve of CC, because the top-10 nodes of them are the same. The number of infected nodes of the proposed method is slightly less than that of LC, but greater than that of others. In Jazz musicians network, the average spreading ability of the proposed method is slightly stronger than CC, LC and DC, the average spreading ability of BC is the worst. In USAir97 network, the number of infected nodes of the proposed method is similar with that of DC, but slightly greater than that of LC and CC. BC has the least number of infected nodes. As for Email network, the proposed method has similar spreading ability with CC, and the spreading ability of them is stronger than LC and DC, Ks has the worst spreading ability. In general, the proposed method and CC have similar spreading ability, which is stronger than other methods.

4) EXPERIMENT 4: COMPARE THE CORRELATION BETWEEN SIX METHODS AND THE ACTUAL SPREADING SITUATION

We take the Kendall’s tau coefficient as a correlation coefficient between the six methods and the actual spreading situation. The ranking list \(\varepsilon\) at \(t = 20\) simulated by SI model is considered as the actual spreading situation, and \(\varepsilon\) is compared with the ranking lists using the six methods. The correlation between different methods and the actual spreading situation with a varying spreading probability from 0 to 0.1. As shown in FIGURE 4, In Karate club network, after spreading probability 0.02, the proposed method and LC have similar \(\tau\) value, and the \(\tau\) value of the proposed method and LC is greater than that of other methods, that is to say, they have the strongest correlation with actual spreading situation. In Jazz musicians network, DC has the highest \(\tau\) value, followed by the proposed method. After spreading probability 0.05, the proposed method and CC have similar \(\tau\) value. In USAir97 network and Email network, the proposed method has the highest \(\tau\) value, which indicates that the proposed method has the strongest correlation with the actual spreading situation. Based on the comparison of the correlation coefficient between the six methods and the actual spreading situation, we can find that the proposed method has the strongest correlation with the actual spreading situation almost in the four networks, and the correlation of BC is the worst.

V. CONCLUSION

Identification and protection of key nodes in complex networks is an effective way to improve the invulnerability and robustness of network. To solve this issue, we propose a novel method to identify key nodes based on global structure, which includes three aspects: the shortest path length, the number of shortest paths and the number of non-shortest paths. On the basis of them, three corresponding influence matrices are established. Since node efficiency can reflect the contribution of one node to the information transmission of the entire network, node efficiency is selected as the initial value of node’s influence on other nodes. We combine the three influence matrices with node efficiency, and establish the comprehensive influence matrix. Then we can obtain the influence of each node in network. To evaluate the performance of the proposed method, we conduct four experiments based on four real datasets, and six classical centralities are applied to the same datasets for comparison. In addition, we simulate the spreading process using SI model with 1000 Monte Carlo runs, and the Kendall’s tau coefficient is also taken as a parameter to compare the performance of different methods. Experimental results show that the proposed method has better performance in identifying key nodes.
ACKNOWLEDGMENT

The authors are grateful to valuable comments and suggestions of the reviewers.

REFERENCES

[1] A.-L. Barabási and E. Bonabeau, “Scale-free networks,” Sci. Amer., vol. 288, no. 5, pp. 60–69, 2003.

[2] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, “Vital nodes identification in complex networks,” Phys. Rev. E, vol. 650, pp. 1–63, Sep 2016.

[3] M. J. Álvarez, Y. Shen, F. M. Giorgi, A. Lachmann, B. B. Ding, B. H. Ye, and A. Califano, “Functional characterization of somatic mutations in cancer using network-based inference of protein activity,” Nature Genet., vol. 48, no. 8, pp. 838–847, Aug. 2016.

[4] B. Huang, X.-Y. Zhao, K. Qi, M. Tang, and Y. Do, “Coloring the complex networks and its application for immunization strategy,” Acta Phys. Sinica, vol. 62, no. 21, 2013, Art. no. 218902.

[5] B. Sidders, A. Karlsson, L. Kitching, R. Torella, P. Karila, and A. Phelan, “Network-based drug discovery: Coupling network pharmacology with phenotypic screening for neuronal excitability,” J. Mol. Biol., vol. 430, no. 18, pp. 3005–3015, Sep. 2018.

[6] W. Yan, D. Zhang, C. Shen, Z. Liang, and G. Hu, “Recent advances on the network models in target-based drug discovery,” Current Topics Med. Chem., vol. 18, no. 13, pp. 1031–1043, Oct. 2018.

[7] Y.-C. Zhang, Y. Liu, H.-F. Zhang, H. Cheng, and F. Xiong, “The research of information dissemination model on online social network,” Acta Phys. Sinica, vol. 60, no. 5, 2011, Art. no. 050501.

[8] Y. R. Gu and L. L. Xia, “The propagation and inhibition of rumors in online social network,” Acta Phys. Sinica, vol. 61, no. 23, 2012, Art. no. 238701.

[9] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, “Breakdown of the Internet under intentional attack,” Phys. Rev. Lett., vol. 86, no. 16, pp. 3682–3685, Jul. 2002.

[10] P. Bonacich, “Factoring and weighting approaches to status scores and clique identification,” J. Math. Sociol., vol. 2, no. 1, pp. 113–120, Jan. 1972.

[11] L. C. Freeman, “Centrality in social networks conceptual clarification,” Social Netw., vol. 1, no. 3, pp. 215–239, Jan. 1978.

[12] M. E. Newman, “A measure of betweenness centrality based on random walks,” Social Netw., vol. 27, no. 1, pp. 39–54, Jan. 2005.

[13] L. C. Freeman, “A new status index derived from sociometric analysis,” Sociometry, vol. 18, no. 1, pp. 39–43, Mar. 1953.

[14] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, “Identifying influential nodes in complex networks,” Phys. A, Statist. Mech. Appl., vol. 391, pp. 1777–1787, Feb. 2012.

[15] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, “Identification of influential spreaders in complex networks,” Nature Phys., vol. 6, no. 11, pp. 888–893, Nov. 2010.

[16] J. Bae and S. Kim, “Identifying and ranking influential spreaders in complex networks by neighborhood coreness,” Phys. A, Stat. Mech. Appl., vol. 395, pp. 549–559, Feb. 2014.

[17] J.-G. Liu, Z.-M. Ren, and Q. Guo, “Ranking the spreading influence in complex networks,” Phys. A, Stat. Mech. Appl., vol. 392, no. 18, pp. 4134–4139, Sep. 2013.

[18] A. Zareie, A. Sheikhalhamedi, and A. Fatemi, “Influential nodes ranking in complex networks: An entropy-based approach,” Chaos, Solitons Fractals, vol. 104, pp. 485–494, Nov. 2017.

[19] Y. Yang, L. Yu, X. Wang, Z. Zhou, Y. Chen, and T. Kou, “A novel method to evaluate node importance in complex networks,” Phys. A, Stat. Mech. Appl., vol. 526, Jul. 2019, Art. no. 121118.

[20] Z. Liu, C. Jiang, J. Wang, and H. Yu, “The node importance in actual complex networks based on a multi-attribute ranking method,” Knowl.-Based Syst., vol. 84, pp. 56–66, Aug. 2015.

[21] L. Fei, H. Mo, and Y. Deng, “A new method to identify influential nodes based on combining of existing centrality measures,” Mod. Phys. Lett. B, vol. 31, no. 26, Sep. 2017, Art. no. 1750243.

[22] S. Wang and J. Zhao, “Multi-attribute integrated measurement of node importance in complex networks,” Chaos, vol. 25, no. 11, Nov. 2015, Art. no. 113105.

[23] T. Zhou, J.-G. Liu, W.-J. Bai, G. Chen, and B.-H. Wang, “Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 74, no. 5, Nov. 2006.

[24] F. Shao and B. Cheng, “Optimal routing strategy based on specifying shortest path,” J. Comput. Commun. Control, vol. 9, no. 5, pp. 602–609, Sep. 2014.

[25] L. C. Freeman, “A set of measures of centrality based on betweenness,” Socioometry, vol. 40, no. 1, pp. 35–41, 1977.

[26] D. A. Griffith and Y. Chun, “Spatial autocorrelation in spatial interactions models: Geographic scale and resolution implications for network resilience and vulnerability,” Netw. Spatial Econ., vol. 15, no. 2, pp. 337–365, Jun. 2015.

[27] W. W. Zachary, “An information flow model for conflict and fission in small groups,” J. Anthropol. Res., vol. 33, no. 4, pp. 452–473, 1977.

[28] P. M. Gleiser and L. Danon, “Community structure in Jazz,” Adv. Complex Syst., vol. 6, no. 4, pp. 565–573, Dec. 2003.

[29] North American Transportation Atlas Data (NORTAD). Accessed: 1997. [Online]. Available: http://vlado.fmf.uni-lj.si/pub/networks/pajek/data/gbhs.htm

[30] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, “Self-similar community structure in a network of human interactions,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 68, Dec. 2003, Art. no. 065103.

[31] M. G. Kendall, “A new measure of rank correlation,” Biometrika, vol. 30, nos. 1–2, pp. 81–93, Jun. 1938.

YUANZHI YANG received the B.S. and M.S. degrees from Air Force Engineering University, Xi’an, China, in 2014 and 2017, respectively, where he is currently pursuing the Ph.D. degree with the Aeronautics Engineering College.

His current research interests include complex networks, failure propagation, and machine learning.

XING WANG is currently a Professor with Air Force Engineering University.

His research interests include electronic countermeasures, machine learning, and artificial intelligence.
YOU CHEN received the Ph.D. degree from Air Force Engineering University, Xi’an, China, in 2011. He is currently a Lecturer with Air Force Engineering University. His research interests include pattern recognition, data mining, machine learning, and artificial intelligence.

MIN HU received the B.S. degree from the School of Liaoning Petroleum Chemical Industry University, China, in 2017. She is currently pursuing the master’s degree with the School of Southwest Petroleum University, China. Her research interests include theoretical calculation of materials and corrosion and protection.