Influencers Identification in Weighted and Directed Networks Based on k-layer Decomposition

Yuecheng Cai* and Wen Zhou

1 School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China

*Corresponding author’s e-mail: cokycyc@163.com

Abstract. In recent years, complex network research has received wide attention from scientific and industrial circles in different academic fields owing to its high application value. In the field of complex network research, identifying influencers has far-reaching implications for various real-world propagation dynamics, such as epidemic outbreaks, and cascading failures. In this paper, we analyse the node-specific attributes of weighted and directed networks and propose two new influencers identification indices for weighted and directed networks. For weighted networks, in k-layer decomposition, weighted degree is used in the algorithm instead of degree as the basis for decomposition. For directed networks, we separately consider the effect of in-degree and out-degree on network decomposition and propose directed k-layer (KLD) indicators that are more suitable for directed networks. Thereafter, a Susceptible-Infected-Recovered (SIR) model and network robustness evaluation methods are used to compare and evaluate the proposed indices on multiple real-world complex network datasets. We found that using the proposed indices resulted in improved node ranking ability for weighted or directed networks.

1. Introduction

It is well known that, in the process of information diffusion, influencers have a great influence on the structure and functioning of networks[1]. Owing to their network structure-related privileges, they make information spread faster and more widely. In recent years, this phenomenon has been widely used in viral marketing[2], computer virus spreading[3], social network analysis[4], epidemic spreading[5], community discovery[6], power grid cascading failure prediction[7], and so on.

Therefore, accurately identifying and locating influencers is critical for designing effective strategies in these different settings[8-10]. Various indices have been proposed for identifying influencers based on centrality-based heuristics that encode the rich structural information of the nodes’ location to distinguish between the influences of nodes in a network, such as degree[11], closeness[12], betweenness[13], k-shell[14], and PageRank[15], just to name a few.

The k-shell decomposition approach[4] was employed by Kitsak et al.[14] in one of their most propounding works. They showed that k-shell decomposition performs better that using the node’s degree for identifying influencers in networks, even though the latter was once considered the simplest and most effective index for this purpose. Nodes with high k-shell values are usually at the core of the network and can initiate large-scale spreading. Therefore, they have stronger spreading capabilities and are regarded as some of the most influential nodes.

However, using the k-shell index for influencers identification has several limitations[16]. For example, the k-shell index is not effective on networks without core structures, such as the Barabási-Albert network[8]. Moreover, the k-shell index lacks good balance between using local and global
structural attributes. For example, highly connected nodes at the periphery of the network could have high k-shell values, suggesting that the k-shell approach uses local structural information and may be incapable of determining the real core of the network.

To overcome the limitations mentioned above, the k-layer index was proposed[17], which is calculated via the k-layer decomposition algorithm. The k-layer index measures the comprehensive distances of nodes from the periphery of the network, taking both global network topology features and local features into consideration in order to measure the coreness of nodes. Yifan[17] compared the k-layer index and the k-shell index in terms of their ability to identify influencers in undirected and unweighted networks and proved that the performance of the k-layer index was better.

The influential nodes identification problem is also called “influencers identification” because its accuracy is usually tested using epidemic spreading models. Epidemic spreading models are a theoretical modeling of how diseases spread in complex networks. They assume that propagation is driven by diffusion processes because transmission occurs from every infected node to all its neighbors at each time step, thus producing a diffusion of the epidemics on the network, such as in the SIR (Susceptible-Infected-Recovered) and SIS (Susceptible-Infected-Susceptible) models[18]. These models usually employ given initially infected nodes to initiate the spreading process in the network. The number of infected nodes is then used for identifying their spreading ability. Furthermore, these models can be used for comparing the indices for influencers identification.

On the one hand, in a weighted network, each node has two basic attributes, namely degree and weight. A node with a high degree in a weighted network may have a small weight, and there may also be nodes with a small degree and high weight. In reality, in many cases, people want to focus on high-weight nodes, which are regarded as influencers in the network. On the other hand, in a directed network, they high out-degree and high in-degree of a given node represent different cases, i.e., high contagiousness and high information volume, respectively. Therefore, it is necessary to study the nodes’ importance in directed networks in terms of both attributes. For the above reasons, the k-layer index does not apply to weighted or directed networks. However, Yifan[17] demonstrated that the k-layer index is superior to the k-shell index, the latter of which is widely recognized for its node ranking performance in undirected complex networks[14]. Thus, based on the k-layer decomposition approach, this paper presents improvements to the k-layer index from two perspectives, with the improved indices exhibiting better performance for node ranking and being more suitable for directed or weighted networks.

2. KW-layer for Weighted Networks

In previous work, we proved through various experiments that the k-layer index has better influencers recognition ability in undirected and unweighted networks than other indices, such as k-shell[17]. However, most networks in the real world are not undirected and unweighted[19]. For example, edge weights in text networks represent statistics on the number of times words co-occur, and edge weights in traffic networks represent the busyness of the lines. Therefore, we considered extending the k-layer index to influencers identification in weighted networks. Figure 1 illustrates the drawbacks of k-layer decomposition when applied to a weighted network.

![Figure 1. A weighted complex network.](image-url)
Assume that the network shown in Figure 1 is a weighted network. The weight of the edge between nodes A and C is set to 3. Then, the weight of the other edges in the network are set to 1. These operations are denoted as $W_{AB} = 1$ and $W_{AC} = 1$.

By decomposing the network via k-layer decomposition, one obtains the k-layer values of nodes A, B, and C, which are respectively denoted as $KL_A = 3$ and $KL_B = KL_C = 0$. In the network, the weighted edge between nodes A and C (which is three times the weight of the other edges) means that the particular nodes associated with this edge are more important than the other nodes at the same location. However, this is not taken into account in k-layer decomposition, resulting in nodes B and C having the same k-layer index, i.e., $KL = 0$.

Therefore, we propose a weighted k-layer decomposition method called kw-layer decomposition. This method still uses the specific steps of k-layer decomposition, but it employs each node's weighted degree instead of its degree. The weighted k-layer index obtained via kw-layer decomposition is named KLW. It takes into account both node degree and node weight and is represents an improvement from the previous method because it employs node weighted degrees instead of node degrees to decompose the network. This algorithm can also be used for undirected and unweighted complex networks.

In the proposed approach, for the two attributes of nodes in weighted networks, we use the weighted degree instead of node degree to perform k-layer decomposition on the network. In a weighted complex network, the weight of a node is the sum of the weights of all the edges associated with that node[20], denoted as $s_i$, whose formula is as shown in the equation (1),

$$s_i = \sum_j w_{ij},$$  \hspace{1cm} (1)

where $j$ is an adjacent node of node $i$ and $w_{ij}$ represents the edge weight between nodes $i$ and $j$. The weighted degree of node $i$ in the network[20] is denoted as $d'_i$, which is calculated as equation (2):

$$d'_i = (d_i a_i b_i)^{\frac{1}{m+1}},$$  \hspace{1cm} (2)

where $a$ and $b$ are the proportions of the node degree and the weighted degree when calculating the weight of nodes, respectively. In our experiments, we only consider cases where $a$ and $b$ are equal to 1 in equation (2); that is, the node degree and the weighted degree have the same importance. These two parameters were then set to explore the impact of different weights on influencers identification in weighted networks in subsequent experiments. Here, the equation of the weighted degree of a node can be simplified as shown in equation (3).

$$d'_i = \sqrt{d_i s_i}$$  \hspace{1cm} (3)

The resulting node weighted degree is then normalized. It can be seen that when these equations are used for an unweighted network, the value of $s_i$ can be regarded as 1. Thus, the weighted degree $d'_i$ is equal to node degree $d$, that is, the k-layer decomposition method is a special case in which the kw-layer decomposition method is applied to an unweighted network. The kw-layer decomposition method consists of four steps:

(i) (Initiation): Set $KLW_i = 0$, where $i = 1, 2, ..., n$ and $k = 1$.

(ii) (Removal): Remove node $t$ if $d'_t \leq k$, and set $KLW_t = \max (KLW_t, KLW_{j_1} + 1, KLW_{j_2} + 1, ..., KLW_{j_s} + 1)$, where $j_1, j_2, ..., j_s$ are the adjacent nodes of node $t$ that were previously removed from the network.

(iii) (Iteration): Repeat Step 2 until there is no node with a weighted degree $d'_t \leq k$.

(iv) (Judgment): If the remaining subgraph is not empty, set $k = k + 1$ and go to Step 2. Otherwise, exit.

The kw-layer is defined as shown in equation (4):

$$KLW_i = \max (KLW_i, KLW_{j_1} + 1, KLW_{j_2} + 1, ..., KLW_{j_s} + 1)$$  \hspace{1cm} (4)
3. KLD index for directed networks

Unlike undirected networks, nodes in directed networks have two attributes, namely out-degree and in-degree, denoted by \( d_{\text{out}} \) and \( d_{\text{in}} \) respectively. They represent the two directions of information propagation and also bear different meanings in real-world networks. For example, in a social network, nodes represent the users of the social network. An edge pointing from node B to A represents the propagation of information from node B to node A. It can be inferred from the above definition that if a node has a high out-degree, a large amount of information is propagated from that node to other nodes, that is, the node has strong infectivity. If a node has a high in-degree, it means that the amount of information received by the node is large.

As reported by Yifan[17], the k-layer index is superior to the k-shell indicator in undirected complex networks. However, neither the k-layer index nor the k-shell index take into account the two attributes of nodes in directed networks. Therefore, for directed networks, we propose k-layer decomposition using these two attributes, namely nodes’ out-degree (denoted as \( d_{\text{out}} \)) and in-degree (denoted as \( d_{\text{in}} \)), in order to obtain the out-degree k-layer index and the in-degree k-layer index, respectively denoted as \( KL_{\text{out}} \) and \( KL_{\text{in}} \).

The specific algorithm for \( KL_{\text{out}} \) decomposition is as follows (the algorithm for \( KL_{\text{in}} \) decomposition is analogous):

(i) (Initiation): Set \( KL_{i}^{\text{out}} = 0 \), where \( i = 1,2, \ldots, n \) and \( k = 1 \).

(ii) (Removal): Remove node \( t \) if \( d_{t}^{\text{out}} \leq k \), and set \( KL_{t}^{\text{out}} = \max (KL_{1}^{\text{out}}, KL_{2}^{\text{out}} + 1, KL_{3}^{\text{out}} + 1, \ldots, KL_{j}^{\text{out}} + 1, J_{1}, J_{2}, \ldots, J_{k} \) are the adjacent nodes of node \( t \) that were previously removed from the network.

(iii) (Iteration): Repeat Step 2 until there is no node with a out-degree \( d_{t}^{\text{out}} \leq k \).

(iv) (Judgment): If the remaining subgraph is not empty, set \( k+1 \) and go to Step 2. Otherwise, exit.

By distinguishing the two attributes of nodes and obtaining respective k-layer indices, we propose a directed k-layer index, denoted as KLD, that more accurately describes the characteristics of the directed network structure. It is defined as shown in equation (5):

\[
KLD_{i} = \frac{mKL_{i}^{\text{out}} + nKL_{i}^{\text{in}}}{m+n},
\]

where \( m \) and \( n \) respectively represent the proportions of \( KL_{i}^{\text{out}} \) and \( KL_{i}^{\text{in}} \). To simplify the calculation, we will round off the resulting KLD. In our experiments, we assume that \( KL_{i}^{\text{out}} \) and \( KL_{i}^{\text{in}} \) account for the same proportion, that is, \( m = n = 1 \). Therefore, we can simplify equation (5) into equation (6).

\[
KLD_{i} = \frac{1}{2}(KL_{i}^{\text{out}} + KL_{i}^{\text{in}}),
\]

4. Results

Six real-world networks were considered in this study: three weighted networks (the first three) and three directed networks (the last three). They can be described as follows:

(i) BibNoun[21]: A weighted network of noun phrases (places and names) in the King James Version of the Bible.

(ii) BusInMbn[22]: A weighted network representing bus routes in Mumbai, India.

(iii) PGP[23]: A weighted network of the giant component of the network of users of the Pretty-Good-Privacy (PGP) algorithm for secure information interchange from 2004.

(iv) Japanese[24]: A directed network of word adjacency in Japanese texts.

(v) EmailRV[25]: An email communication network from the University Rovira i Virgili in Tarragona in the south of Catalonia, Spain.

(vi) OpenFlights[26]: A directed network containing flights between airports of the world extracted from Openflights.org data.

The network data used for the experiments are public standard datasets (see references for details). Detailed information about the networks is presented in Table 1.
Table 1. Properties of the real-world networks considered in this work: number of nodes ($N$), number of edges ($E$), average degree ($\bar{d}$), clustering coefficient ($C$), and epidemic threshold ($\beta_C$).

| Network    | N   | E    | $\bar{d}$ | C   | $\beta_C$ |
|------------|-----|------|-----------|-----|-----------|
| BibNoun    | 1773| 9131 | 5.15      | 0.36| 0.024     |
| BusInMbn   | 2266| 3042 | 1.342     | 0.075| 0.445     |
| PGP        | 10680| 24316| 2.277     | 0.133| 0.056     |
| Japanese   | 3177| 8300 | 3.070     | 0.138| 0.008     |
| EmailRV    | 1133| 10902| 9.622     | 0.220| 0.027     |
| OpenFlights| 2939| 30501| 10.378    | 0.435| 0.008     |

4.1. kw-layer for weighted networks

4.1.1. Results Based on an Epidemic Spreading Dynamics Model. To compare the effectiveness of the different indices, their Kendall correlation coefficient $\tau$ (introduced in detail in the Methods section) values were calculated using the simulation propagation results obtained via an SIR model. The evaluation method is also described in detail in the Methods section. In the simulations, the infection coefficient $\beta$ was set to slightly higher than the epidemic threshold $\beta_C$, that is, $\beta = 1.5\beta_C[19]$. The values of the epidemic threshold $\beta_C$, for the different considered networks are indicated in column 6 of Table 1.

We horizontally compared the ability of the KLW index and the k-layer index to identify influencers on weighted networks. As shown in Table 2, for the three real-world weighted networks (BibNoun, BusInMbn, and PGP), their Kendall coefficients $\tau_{KLW}$ are all greater than $\tau_{KL}$. This indicates that, compared with the node sequences of the k-layer approach, the node sequences of the KLW index approach are more similar to those of the SIR model. Because we used the node sequences of the SIR model as the standard, we can conclude from this experiment that the KLW index is better than the k-layer index in terms of its ability to sort nodes in weighted networks.

Table 2. A Kendall correlation coefficients $\tau$, which show the correlation between the ordering of nodes obtained using different ranking metrics and the ordering of nodes obtained via SIR simulations. KL indicates k-layer index.

| Network    | $\beta = 1.5\beta_C$ | $\tau_{KL}$ | $\tau_{KLW}$ |
|------------|----------------------|-------------|--------------|
| BibNoun    | 0.036                | 0.378       | 0.386        |
| BusInMbn   | 0.667                | 0.049       | 0.179        |
| PGP        | 0.084                | 0.379       | 0.385        |

4.1.2. Results Based on Network Robustness. Here, network robustness evaluation criteria is used to evaluate five indices, namely weighted degree[20], k-shell index[14], k-layer index[17], kw-shell index[20], and KLW index. A node removal strategy was applied to the three weighted networks considered in this study to evaluate the ability of the KLW and the other indices to identify influencers. The network robustness $R$ for each index (see the Methods section for calculations) were calculated and are displayed in Table 3. The smaller the calculated network robustness $R$ is, the better the node ranking ability of the corresponding index is. It can be seen from Table 3 that in all three networks, the network robustness $R$ obtained by removing the nodes according to the node sequence of KLW is smaller than that obtained using the other four indices. Thus, we can conclude that the KLW index represents an improvement in terms of its ability to identify influencers in weighted networks.
Table 3. Network Robustness $R$ of five sequences after removing nodes from the network according to five indices: weighted degree, $k$-shell, $k$-layer, weighted $k$-shell, and KLW are represented in order.

| Network  | $R_{WD}$ | $R_{KS}$ | $R_{KL}$ | $R_{KSW}$ | $R_{KLW}$ |
|----------|----------|----------|----------|-----------|-----------|
| BibNoun  | 0.145    | 0.176    | 0.144    | 0.138     | 0.137     |
| BusInMbn | 0.079    | 0.268    | 0.120    | 0.090     | 0.080     |
| PGP      | 0.053    | 0.116    | 0.070    | 0.053     | 0.050     |

In Figure 2, the number of removed nodes was used as the abscissa and the number of nodes in the giant component was used as the ordinate. As can be observed from Figure 2 and focusing on the BusInMbn dataset, most of the curve for the $kw$-layer index appears to be below the other curves in the chart. This means that removing nodes from the network according to the node sequence obtained using the KLW index has the greatest impact on the structure of the network.

4.2. KLD Index for Directed Networks

4.2.1. Results Based on an Epidemic Spreading Dynamics Model. In this evaluation experiment, we compared five indices, namely the out-degree, $k$-shell, $k$-layer, KLD, and PageRank (detailed in the Methods section) indices. The correlation coefficients $\tau$ of the node sequence obtained by calculating the SIR model and of the node sequence calculated using the abovementioned indices are displayed in Table 4. As shown in Table 4, for the EmailRV and OpenFlights datasets, the value of $\tau_{KLD}$ is the largest (marked in bold), which indicates that for these node ranking indices, the node sequence obtained using the KLD index is most similar to the node sequence of the SIR model. This means that the KLD index performs best for two of the considered datasets. For the Japanese network, the KLD index performed slightly worse than the $k$-layer index but still outperformed the others. It can be seen that the KLD index maintains excellent and even slightly improved performance for influencers identification in directed networks.
Table 4. A Kendall correlation coefficients $\tau$, which show the correlation between the ordering of nodes obtained using different ranking metrics and the ordering of nodes obtained via SIR simulations.

| Network | $\beta = 1.5\beta_C$ | $\tau_{d_{out}}$ | $\tau_{KS}$ | $\tau_{KL}$ | $\tau_{KLD}$ | $\tau_{PageRank}$ |
|---------|---------------------|-----------------|-------------|-------------|--------------|------------------|
| Japanese | 0.012               | 0.174           | 0.188       | **0.233**   | 0.224        | 0.179            |
| EmailRV | 0.041               | 0.368           | 0.359       | 0.364       | **0.379**    | 0.352            |
| OpenFlights | 0.012             | 0.327           | 0.323       | 0.346       | **0.350**    | 0.289            |

4.2.2. Results Based on Network Robustness. We removed nodes from networks in order according to the node sequences obtained after calculating the five indices, namely the out-degree, k-shell, k-layer, KLD, and PageRank indices. Their respective network robustness $R$ values are shown in Table 5. For the Japanese network, although $\tau_{KLD}$ was slightly worse than $\tau_{KL}$ in the SIR model evaluation, the performance of the KLD index was the best in this network robustness experiment, and $R_{KL}$ only ranked fourth for all networks. As for the EmailRV and OpenFlights networks, $R_{KLD}$ ranked second, just slightly worse than $R_{PageRank}$. However, in the SIR model experiment, the KLD index yielded the best results, whereas PageRank only ranked fifth. Therefore, based on the combined results of the two experiments, we believe that the KLD index represents an improvement in terms of its node ranking ability compared with the k-layer index for directed networks and has certain advantages in terms of influencers identification performance compared with the other four indices.

Table 5. Network Robustness $R$ of five sequences after removing nodes from the network according to five indices: out-degree, k-shell index, k-layer index, KLD index, and PageRank are represented in order.

| Network | $R_{d_{out}}$ | $R_{KS}$ | $R_{KL}$ | $R_{KLD}$ | $R_{PageRank}$ |
|---------|---------------|----------|----------|-----------|----------------|
| Japanese | 0.00386       | 0.00436  | 0.00431  | **0.00352** | 0.00405       |
| EmailRV | 0.2519        | 0.2937   | 0.2728   | 0.2479    | **0.2396**    |
| OpenFlights | 0.0812     | 0.1089   | 0.0858   | 0.0811    | **0.0666**    |

For Figure 3, we used the Japanese network as an example; the number of removed nodes was used as the abscissa and the number of nodes in the giant component was used as the ordinate. We can observe that most of the red curve, which represents the KLD index, is located below the other curves, indicating that removing nodes from the network using the node ordering obtained with the KLD index has the greatest impact on network robustness.

Figure 3. Number of nodes in the giant component versus number of nodes removed in descending order according to each ranking index.
5. Conclusion
This paper extends the k-layer index in two directions and proposes two new indices with better influencers identification ability on both directed networks and weighted networks. For weighted networks, by analysing the two attributes of nodes, namely node degree and weight, the network can be decomposed using the weighted degree of nodes; the KLW index, which is suitable for weighted networks, is thus proposed. This index takes node weight into account and is more consistent with the topological characteristics of weighted networks than the k-layer index. For directed networks, we propose the KLD index by distinguishing the influence of each node's out-degree and in-degree on node importance based on k-layer decomposition. The proposed index is more suitable for influencers identification in directed networks compared with four other indices.

To compare the abilities of different indices for node influence ranking, the Kendall correlation coefficient values for the different ranking indices relative to the SIR model were calculated. The impact on the robustness of networks resulting from the removal of nodes in the order of the rankings obtained using three indices was also observed, which was quantified via the number of nodes that remain in the giant component of the network. We found that the two proposed indices generally produced better assessments of the impact of nodes compared with the other indices.

In future work, based on k-layer decomposition, we will consider combining the two indices proposed in this paper to explore common indices that can be applied to directed and weighted networks. In addition, we will also consider improving the k-layer index from the perspective of neighbour nodes to further improve its performance for influencers identification.

6. Methods

6.1. Evaluation Methods
This section first reviews several centrality indices, which will be used for comparison with the proposed KLW and KLD indices. The evaluation criteria are then presented, consisting in two different evaluation methods that ensure that the different indices are compared in every possible way.

In a network $G = (V,E)$, where $V$ and $E$ represent the nodes and edges in the network, respectively, the degree of node $i$, denoted by $d_i$, is defined as the number of its directly connected neighbours. The k-shell coreness of node $i$ measured via k-shell decomposition is denoted by $K_S_i$. Coreness $K_S_i$ indicates that node $i$ belongs to a k-shell but not to any $(k+1)$-shell[14].

The weighted k-shell index (denoted as kw-shell) is obtained via k-shell decomposition of the network using node weighted degrees[20]. This process is analogous to the k-shell decomposition process[14].

In the k-shell decomposition process, all nodes with degree $k \leq 1$ and their links are first removed. This step is repeated until there are no nodes with degree $k \leq 1$ in the network. All the nodes that have been removed at this point are those that constitute the 1-shell. The degree of the remaining nodes in the network is referred to as the residual degree.

The above process is repeated to further decompose the network into 2-shell, 3-shell, etc., until there are no more nodes in the network. The coreness $K_S$ of a node is thus equal to the $k$ of the k-shell that it belongs to and is used to describe the topological location of the node in the network.

The PageRank approach supposes that the importance of a node is determined by the quantity and quality of the nodes connected to it[15]. Initially, each node is assigned a unit PageRank value (denoted as $PR$). Each node then evenly distributes their $PR$ value to its neighbours along its outgoing edges until the $PR$ values of all nodes reaches a steady state.

When evaluating indices for ranking the importance of nodes in a network, the focus is usually placed on whether the index accurately reflects the position of the nodes in the network. However, it is difficult to objectively determine the position of a node in a large-scale network.

There are two common evaluation criteria. The first one is based on an epidemic spreading dynamics model, which is used to examine the infectious ability of the node on other nodes in the network[18].
Infectious source nodes that are ranked higher in the infectious disease model are considered more susceptible to other nodes in the network.

The second evaluation criterion is based on network robustness and is used to sort nodes based on their ability to influence the network’s structures and functions. According to Cohen et al.[27], if the removal of a node or group of nodes would significantly reduce the giant component of the network, that node or nodes can be considered to be important for the network. A ranking index is considered good when the sequence of nodes sorted by importance that is obtained using it can be used to attack the network more effectively.

6.2. Evaluation Criteria Based on an Epidemic Spreading Dynamics Model
This section first reviews several centrality indices, which will be used for comparison with the proposed KLW and KLD indices. The evaluation criteria are then presented, consisting in two different evaluation methods that ensure that the different indices are compared in every possible way.

In the present study, to evaluate the performance of different node importance ranking indices, an SIR model was used to simulate the diffusion effects of the nodes in different networks[28].

An initially infected node $i$ was set in the SIR model and all the other nodes were set to be susceptible at the initial stage. Each infected node then subsequently infected its susceptible neighbours with an infection probability $\beta$ and was restored to a steady state with a recovery probability $\mu$. This infection process was repeated until there were no infected nodes in the network. The number of nodes in the restored state at the end of the epidemic infection was the evaluation index value of the infection ability of the initially infected node $i$.

In the employed SIR model, the recovery probability $\mu$ was set to 1, indicating that a node was restored to a steady state immediately after being infected. The infection probability $\beta$ should not be set too low nor too high. If $\beta$ were too low, it could prevent the epidemic from successfully propagating, and the communication capability of nodes would be impossible to measurable. Conversely, if $\beta$ were too high, all the nodes in the network could be infected and it would thus become difficult to distinguish between the diffusion capabilities of different nodes. According to Lin et al.[29], when $\beta$ approaches the epidemic threshold $\beta_C$, the central behavior of the eigenvector is enhanced[30]. The epidemic threshold $\beta_C$ is given by equation (7):

$$\beta_C = \frac{d}{\bar{d}^2 - \bar{d}}, \quad (7)$$

where $d$ is the degree of the node, $\bar{d}$ denotes the average degree of the nodes in the network, and $\bar{d}^2$ denotes the average of the squares of the degrees of all nodes.

To quantify the correctness of each of the ranking methods, Kendall’s rank correlation coefficient $\tau$ was used as an indicator[31]. For two sequences $X$ and $Y$, if the elements in each sequence are unique, $\tau$ can be calculated using equation (8)[32]:

$$\tau = \frac{C - D}{0.5N(N-1)}, \quad (8)$$

where $C$ is the number of pairs of consistent elements in $X$ and $Y$ (two elements form a pair), $D$ is the number of pairs of inconsistent elements in $X$ and $Y$, and $N$ is the sequence length. This indicator quantifies the similarity between the ordering of the elements in the two sequences. The higher the correlation coefficient $\tau$ is, the better the ranking index is.

6.3. Evaluation Criteria Based on Network Robustness
Intentional attacks can have devastating effects on the connectivity of a complex network. An evaluation criterion based on network robustness was used to assess the effects of removing a set of nodes on the network structure and functions[19]. The greater the resulting impact is, the more important the removed node is.

A node or group of nodes can be evaluated to determine whether they are critical to the network by observing the reduction of the giant component of the network caused by their removal. The importance
of the core nodes of a network can also be quantified by the number of connected components that remain after their removal. According to the node removal policy employed, nodes are generally removed in descending order according to their ranking indices. In the present study, our focus was on the damage caused to network stability or robustness. That is, a node or group of nodes were considered to be influencers in the network if their removal significantly reduced network robustness.

The robustness of a network can be characterized as $R$, which is calculated using equation (9):

$$
R = \frac{1}{2} \sum_{i=1}^{n} \sigma \left( \frac{i}{n} \right),
$$

where $\sigma \left( \frac{i}{n} \right)$ is the proportion of the number of nodes belonging to the giant component of the network after the removal of $\frac{i}{n}$ of the nodes and, $i$ is the number of nodes removed, and $n$ is the number of nodes in the network. In the evaluation experiment, we calculated the network robustness $R$ corresponding to different orders of removal obtained according to different node ranking indices. It is clear that the lower the calculated network robustness $R$ is (after the removal of nodes), the better the performance of the corresponding node importance index is.

Acknowledgments
We would like to thank Editage [www.editage.cn] for English language editing.

References
[1] Kempe, D., Kleinberg, J., Tardos, A. (2003) Maximizing the spread of influence through a social network. ACM SIGKDD Int. Conf. on Knowl. Discov. and Data Min., 137–146.
[2] Leskovec, J., Adamic, L. A., Huberman, B. A. (2007) The dynamics of viral marketing. ACM Trans. Web 1.
[3] Pastor-Satorras, R., Vespignani, A. (2007) Evolution and structure of the Internet: A statistical physics approach. Cambridge University Press, London.
[4] Liu, S. et al. (2015) Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach. Inf. Sci. 306, 34–52.
[5] Pastor-Satorras, R., Castellano, C., Van Mieghem, P., Vespignani, A. (2015) Epidemic processes in complex networks. Rev. Mod. Phys. 87, 925–979.
[6] Bao, W., Michailidis, G. (2018) Core community structure recovery and phase transition detection in temporally evolving networks. Sci. rep. 8, 12938.
[7] Yang, Y., Nishikawa, T., Motter, A. E. (2017) Small vulnerable sets determine large network cascades in power grids. Science 358, eaan3184.
[8] Barabási, A.-L., Albert, R. (1999) Emergence of Scaling in Random Networks[J]. Science, 286.
[9] Pei, S., Makse, H. A. (2013) Spreading dynamics in complex networks. J. Stat. Mech. Theory Exp. 2013, P12002.
[10] Yu, E.-Y., Chen, D.-B., Zhao, J.-Y. (2018) Identifying critical edges in complex networks. Sci. Rep. 8, 14469.
[11] Freeman, L. C. (1978) Centrality in social networks conceptual clarification. Soc. Netw. 1, 215–239.
[12] Sabidussi, G. (1966) The centrality index of a graph. Psychometrika 31, 581–603.
[13] Freeman, L. C. (1977) A set of measures of centrality based on betweenness. Sociometry 40, 35–41.
[14] Kitsak, M. et al. (2010) Identification of influential spreaders in complex networks. Nat. Phys. 6, 888–893.
[15] Sehgal, U., Kaur, K., Kumar, P. (2009) The anatomy of a large-scale hyper textual web search engine. Int. Conf. on Comput. Electr. Eng. 2, 491–495.
[16] Zeng, A., Zhang, C.-J. (2012) Ranking spreaders by decomposing complex networks. Phys. Lett. A 377, 1031–1035.
[17] Yifan, J. (2017) Research on New Method of Key Node Identification and Link Prediction on Complex Networks[D]. Shanghai University.

[18] Barrat, A., Barthlemy, M., Vespignani, A. (2008) Dynamical Processes on Complex Networks. Cambridge University Press, London.

[19] Lü, L. et al. (2016) Vital nodes identification in complex networks. Phys. Rep. 650, 1–63.

[20] Garas, A., Schweitzer, F., Havlin, S. (2012) A k-shell decomposition method for weighted networks. New J. of Physics. 14(8):083030.

[21] Harrison, C. Bible Cross-References (2014) (http://www.chrisharrison.net/index.php/Visualizations/BibleViz)

[22] Chatterjee, A., Manohar, M., Ramadurai, G. (2015) Statistical analysis of bus networks in india. Plos. One. 11.

[23] Rossi, R. A., Ahmed, N. K. (2015) The Network Data Repository with Interactive Graph Analytics and Visualization. Proc. of the 29th AAAI Conf. on Artif. Intell. AAAI Press.

[24] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., Barabási, A.-L. (2000) The large-scale organization of metabolic networks. Nature 407, 651–4.

[25] Guimerà, R., Danon, L., Díaz Guilera, A., Giralt, F., Arenas, A. (2004) Self-similar community structure in a network of human interactions. Phys. Rev. E. Stat. Nonlin. Soft Matter Phys., 68 (6 Pt. 2), 065103.

[26] Opsahl, T., Agneessens, F., Skvoretz, J. (2010) Node centrality in weighted networks: generalizing degree and shortest paths. Soc. Netw., 32(3), 245–251.

[27] Cohen, R., Erez, K., ben Avraham, D., Havlin, S. (2001) Breakdown of the internet under intentional attack. Phys. Rev. Lett. 86, 3682–5.

[28] Newman, M. E. (2002) Spread of epidemic disease on networks. Phys. Review. E Stat. Nonlin. Soft Matter Phys. 66, 016128.

[29] Lin, J.-H., Guo, Q., Liu, J.-G., Zhou, T. (2015) Locating influential nodes via dynamics-sensitive centrality. Sci. Rep. 6, 032812.

[30] Castellano, C., Pastor-Satorras, R. (2010) Thresholds for epidemic spreading in networks. Phys. Rev. Lett 105, 218701, DOI: 10.1103/PhysRevLett.105.218701.

[31] Kendall, M. (1945) The treatment of ties in rank problems. Biometrika 33, 239–51, DOI: 10.1093/biomet/33.3.239.

[32] Kendall, M. (1938) A new measure of rank correlation. Biometrika 30, 81–93, DOI: 10.1093/biomet/30.1-2.81.