

Node Importance Ranking of Complex Network based on Degree and Network Density

Hui Xu^{a,b}, Jianpei Zhang^{a,*}, Jing Yang^a, and Lijun Lun^c

^aCollege of Computer Science and Technology, Harbin Engineering University, Harbin, 150001, China

^bLibrary, Heilongjiang University of Chinese Medicine, Harbin, 150040, China

^cCollege of Computer Science and Information Engineering, Harbin Normal University, Harbin, 150025, China

Abstract

Node importance ranking of complex networks is of great significance to the study of network robustness. The classical centrality measure degree can reflect the number of neighbors of a node, but it ignores the information between its neighbors. In order to mine the important nodes in the network accurately and efficiently, a method of ranking the node importance of complex networks based on multi-attribute evaluation and node deletion is proposed in this paper. Based on the degree attributes of the target node and its neighbors, this method introduces two attributes, which are the local network density centered on the target node and the assortativity coefficient. It takes into account the characteristics of the scale, tightness, and topology of the local area network where the node and its neighbors are located. This paper conducts deliberate attack experiments on four real networks. Through a comparison between the experimental results of the maximal connected coefficient and network efficiency, our approach is proven to be valid and feasible.

Keywords: node importance; robustness; degree; network density

(Submitted on November 10, 2018; Revised on December 12, 2018; Accepted on January 8, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

Many complex systems in the objective world can be described as complex networks. There are numerous examples of complex networks in different subjects, such as biology, sociology, technology, and economics. Although the nodes and edges of complex networks represent different things and relationships in different subject areas, they show some common structural characteristics, such as small-world characteristics [1], scale-free properties [2], robustness to random attacks, vulnerability to malicious attacks, and so on. This makes some special nodes in the network have a great impact on the structure and function of the network. We call these nodes important nodes. In complex networks, a small number of key nodes play an important role in the overall performance of the network. Research shows that the network will be paralyzed [3] when 5% of the important nodes in a scale-free network are attacked. Therefore, it is instructive to evaluate the importance of network nodes and effectively excavate important nodes according to their quantitative data, which are of theoretical significance to the complex network applications in reality. For example, inhibiting the spread of the virus [4-5], locating the leaders of terrorist organizations, avoiding the cascade failure of power networks [6-7], determining the sorting of search results of search engines, and mining community centers in complex network community structure involve the calculation of node importance assessment.

In recent years, how to identify important nodes has been a popular research topic in the study of the structural characteristics of complex networks [8], from which many classical node importance ranking algorithms have been derived [9].

A ranking method is based on the neighborhood of the nodes. This is the simplest and most intuitive method. Degree [10] is a simple and effective local algorithm, but it does not take into account the location of the node and the information of its neighbors. Taking full account of the first-order neighbor and second-order neighbor information of the target node,

* Corresponding author.

E-mail address: zhangjianpei@hrbeu.edu.cn

Chen et al. [11] expanded the coverage range of the node domain and proposed a semi-local centrality method based on neighborhood information to identify the most influential nodes in the undirected network. It achieves a good balance between accuracy and time complexity of the algorithm. Liu et al. [12] proposed a weight degree centrality method that consists of the degree of a node and its neighborhoods, and this method worked well in most experiments. Kitsak et al. [13] believed that the node importance was related to the location of the network, and they proposed a k-shell decomposition method. The computational complexity of this method is low, but the ranking results are too coarse-grained, which causes small discrimination of nodes. This method is not suitable for tree diagrams, regular networks, and BA networks [2]. Many scholars have improved and extended the k-shell decomposition method to overcome its limitations. Zeng et al. [14] proposed the mixed degree decomposition, which fully considered the residual degree and the exhausted degree. Bae and Kim [15] proposed a new node importance metric, extended neighborhood coreness centrality based on the k-shell values of the node's neighbor set, to evaluate the propagation impact of the node. Drawing on the idea of classical gravity formula, Ma et al. [16] proposed a gravity model, which regarded the k-shell value of the node as its mass and the shortest distance between two nodes in the network as their distance, to evaluate the node importance. Based on the definition of entropy in information theory, Liu et al. [17-18] proposed the connection entropy to measure the diversity of network shell connection. This method effectively demonstrates the fact that the k-shell method may identify false cores in the application process due to the local network being too close.

A ranking method based on path: generally, this method is closely related to the path of the network. It investigates the control power of node degree information flow, but the time complexity is relatively high. The betweenness centrality [19] describes the ability to control the transmission of information along the shortest path between nodes and other pairs of nodes in the network. The higher the betweenness of the node, the greater its importance. Closeness centrality [20] eliminates the interference of special values by calculating the average distance between nodes and all other nodes in the network. Katz centrality [21] uses matrix inversion to judge the node importance, which is mainly used in networks with small scale and few loops. Communicability [22] considers not only the shortest path, but also the effect of all paths on nodes, and it assigns smaller weight to longer paths.

A ranking method based on eigenvector: this method considers not only the number of node neighbors, but also the influence of its quality on the importance of nodes. Yang et al. [23] proposed a time series network construction method based on similarity-based supra-adjacency matrix (SSAM). Using the eigenvector centrality [24], the order of importance of nodes at each time horizon is obtained, and the trajectory of node importance changes with time is obtained. PageRank [25] and LeaderRank [26] are famous web page ranking algorithms in the field of search engine. Gu et al. [27] proposed the node similarity index as a measure of the local importance of the node. Based on LeaderRank and considering the local importance and global importance of the node, the SRank algorithm was proposed.

A ranking method based on node removal and shrinkage: the remarkable feature of this method is that in the process of ranking important nodes, the structure of the network will be dynamically changing. The node importance can be measured by the destructive effect after removing nodes on the network. With further study of complex networks, more and more scholars have realized that the evaluation algorithm of node importance, which fuses more information, can more reasonably evaluate the node importance.

In this paper, motivated by authors in these previous studies, a new node importance metric method is proposed based on the degree and the density of the local area network. This method not only considers the neighbor information of the node and the tightness between its neighbors, but also takes into account the network topology of different networks and introduces the assortativity coefficient of the network. The method proposed in this paper combines the local information of nodes and the topology of the network to quantitatively and comprehensively describe the node importance in complex networks. The main contributions of this paper include: (1) taking the target node as the center, a local area network is constructed together with its neighbors. The density of the local area network is proposed and the degree of tightness between the nodes is discussed. (2) Based on the degree, local area network density, and assortativity coefficient, a metric index of important nodes for multi-attribute evaluation is proposed, which is suitable for complex networks with different network topologies. (3) Deliberate attack simulation experiments on four real and open social network data sets are to testify that our method yields better performance of identifying important nodes than many previous methods.

2. Methods

Consider an undirected network $G = (N, M)$ with N nodes and M edges. Given the adjacency matrix $A = (a_{ij})_{N \times N}$ of the network G , the degree of node i can be expressed as:

$$k_i = \sum_{j=1}^N a_{ij} \quad (1)$$

In Equation (1), node i is connected to node j and the a_{ij} value is one; otherwise, it is zero. The sum of neighborhood degrees of node i can be represented as:

$$K_i = \sum_{v \in \Gamma(i)} k_v \quad (2)$$

In Equation (2), $\Gamma(i)$ denotes the set of neighbor nodes of node i . The degree index reflects the ability of establishing a direct connection between nodes and their neighbor nodes, but it does not consider the characteristics of the tightness between neighbors. Therefore, it cannot reflect the interaction between neighbor nodes in the calculation, and the calculation results are not accurate enough.

Suppose we build a local area network with node i as the center and the shortest distance from node i to other nodes as the radius. The scale of the local area network varies with the radius. The radius of the local area network in this paper takes a value of one. Taking node B in Figure 1 as an example, $\{B, A, C, D\}$ constitutes a local area network with a radius of one centered on B . We use the ratio between the number of edges between the target node i and other nodes and the maximum number of possible edges in the local area network centered on node i as the local area network density. The density of the local area network of i -centric can be expressed as:

$$LD_i = \frac{2l_i}{n_i(n_i - 1)} \quad (3)$$

In Equation (3), l_i indicates the number of edges actually presented between node i and other nodes in the i -centric local area network. n_i indicates the number of nodes in the local area network. To some extent, LD_i represents the complexity and the number of relationships in the network. The closer the actual number of edges to the number of edges of the complete graph, the greater the density of the network and the smaller the opposite. The greater the network density, the closer the relationship between nodes in the network and the higher the closure between neighboring nodes. Such nodes are disadvantaged in the competition because they do not easily acquire new relational resources. From the perspective of complex networks, the network density can be used to estimate the node importance. That is, the smaller the network density, the more favorable it is to obtain new relational resources and the greater the importance of nodes. It can be seen from Figure 1 that the density of the local area network centered on node A is smaller than the density of the local area network centered on node B . This shows that node A has a variety of outward connections, and the closure between nodes in the local area network is low, making it easier to acquire new resources. The density of the local area network is susceptible to the scale of the network and the nature of the relationships in the network, making it impossible to objectively measure the density difference between different structural networks. Therefore, in the specific use process, if it can be expressed together with the scale of the local area network and the connection tendency between network nodes, the reference value of the network density will be more obvious.

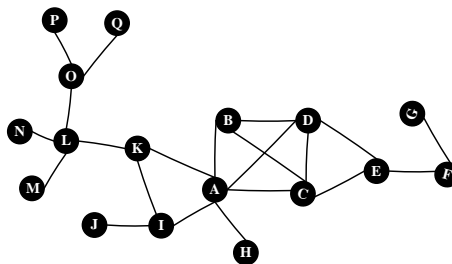


Figure 1. Network with seventeen nodes

Based on the above-mentioned characteristics of the scale, tightness, and topology of the local area network, this paper proposes a new node importance metric (labeled as KLD) and uses the homogenization function [28] $\sigma(x) = x / \sqrt{\sum x^2}$ to deal with the degree and the density of the local area network.

$$KLD_i = \left(\frac{(k_i + K_i)}{\sqrt{(k_i + K_i)^2 + \sum_{u \in \Gamma(i)} (k_u + K_u)^2}} \right)^{|r|} \left(\frac{LD_i}{\sqrt{LD_i^2 + \sum_{w \in \Gamma(i)} LD_w^2}} \right)^{(-1)} \quad (4)$$

In Equation (4), $(k_i + K_i)$ represents the sum of the degree of node i and its neighbor degree. LD_i represents the network density of a local area network with a radius of one centered on i . r denotes the assortativity coefficient of the network, which is used to examine whether nodes with similar degrees tend to be interconnected. The value range of r is $[-1, 1]$. If the large-degree nodes tend to connect with other large-degree nodes, then the degree of the network is positively correlated. That is, the network is assortativity and the value of r is positive. If the large-degree nodes tend to connect with small-degree nodes, then the degree of the network is negatively correlated. That is, the network is heterogeneous and the value of r is negative. The KLD index not only reflects the scale of the target node neighbors, but also the tightness of the local area network and the connection tendency between network nodes.

Taking node A and node B in Figure 1 as an example, the two nodes' neighbor sets are $\Gamma(A)=\{B, C, D, I, K, H\}$ and $\Gamma(B)=\{A, C, D\}$. The sum of neighborhood degrees of them are $K_A = k_B + k_C + k_D + k_I + k_K + k_H = 18$ and $K_B = k_A + k_C + k_D = 14$. The density of the local area network of A -centric and B -centric are $LD_A = 0.2857$ and $LD_B = 0.5000$. Therefore, their KLD values are $KLD_A = 4.8939$ and $KLD_B = 1.5145$. It is not difficult to see that compared with node B , the density of the local area network centered on node A is small, the external connection is diverse, it is easy to acquire new resources, and its importance is relatively higher.

3. Data Experiments and Results Analysis

3.1. Datasets

In order to test the validity and accuracy of the KLD index to evaluate important nodes, we selected four real and open social network data sets with different network topology characteristics for comparative analysis. Zachary's karate club is a social network of friendship between 34 members of a karate club at a US university in the 1970s. Dolphin social network [29] is an undirected social network of frequent associations between 62 dolphins in a community living off doubtful sound. Books about US politics are a network of books about US politics published around the time of the 2004 presidential election and sold by the online bookseller Amazon. Neural network [30] is a network representing the neural network of *Elegans*. Table 1 shows the basic structural characteristics of the four networks. It can be seen from Table 1 that the assortativity coefficients of the four networks are all less than zero, indicating that the low-degree nodes are more likely to be connected with the large-degree nodes, which is consistent with the structural characteristics of the social network.

Table 1. Basic structural characteristics of the four real networks studied in this work
 N and E are the number of nodes and edges, respectively. $\langle k \rangle$ is the average degree, k_{\max} is the maximum degree, r is the degree assortativity.

Network	N	E	$\langle k \rangle$	k_{\max}	r
Karate network	34	78	4.5882	17	-0.4756
Dolphin social network	62	159	5.1290	12	-0.0436
Books about US politics	105	441	8.4000	25	-0.1279
Neural network	297	2359	14.4646	134	-0.1632

3.2. Evaluation Criteria

The node importance can be measured not only by the ability of the node to disseminate information in the network, but also by the destructiveness of the network connectivity after the node is removed. Studies have shown that complex networks are highly resistant to random attacks, but they are very vulnerable to deliberate attacks, and 5%-10% of important nodes may cause the entire network to falter [31]. Therefore, this paper evaluates node importance in the network by deleting nodes and judging the quality of network connectivity. That is, after the node is deleted, the greater the network connectivity changes, and the more important the deleted node is in the network.

3.2.1. Maximal Connected Coefficient

The maximal connected coefficient μ sorts the nodes according to the importance evaluation method from large to small, and it observes the influence of removing some nodes on the network maximal connected subgraph [32-33]. The calculation formula is as follows:

$$\mu = \frac{F}{N} \quad (5)$$

In Equation (5), N is the total number of nodes in the network and F is the number of nodes of the maximal connected subgraph after removing a part of the nodes. The tendency for the scale of the maximal connected subgraph to become smaller as the node is removed becomes more obvious, indicating that the effect of attacking the network by this method is better.

3.2.2. Network Efficiency

Network efficiency [34] is an important index to judge the network connectivity. The worse the network connectivity and the lower the network efficiency, the more important the deleted node sets. The specific formula of network efficiency is as follows:

$$\delta = \frac{1}{N(N-1)} \sum_{i \neq j \in G} (1/d_{ij}) \quad (6)$$

In Equation (6), d_{ij} represents the shortest path between two nodes in the network. The range of δ is $[0, 1]$. $\delta = 1$ indicates the best network connectivity, and $\delta = 0$ indicates that the network is composed of isolated nodes. When there are multiple paths between node i and node j , we interrupt some of the paths, which may make the shortest path d_{ij} between these two nodes become longer. The average path length of the entire network will also become longer, resulting in network connectivity getting worse. δ is normalized to its possible largest value $N(N-1)$ for a completely connected graph having $N(N-1)/2$ edges.

We select the top-ranking important nodes with a certain proportion p to conduct a deliberate attack simulation experiment on the network. The range of p is $[0, 1]$. We calculate the descending ratio of network efficiency before and after attacking to quantitatively describe the accuracy of various node importance indices. Assuming that the network is not attacked, the network efficiency is δ_0 , and then the network efficiency is δ after deleting a certain proportion p of top-ranking important nodes. The descending ratio of network efficiency is expressed as:

$$e = 1 - \frac{\delta}{\delta_0} \quad (7)$$

Where the range of e is $[0, 1]$. When $e = 1$, it indicates that the network efficiency drops to zero after the network is attacked. That is, the network consists of isolated nodes. When $e = 0$, this means that the network efficiency of the entire network has not changed after the network is attacked. It can be seen from Equation (7) that the larger the e value is, the worse the network efficiency becomes after deleting the nodes with a certain proportion p . Therefore, the node importance can be accurately measured by using this method.

3.3. Contrast Indices

In the following experiments, the *KLD* index will be compared with six evaluation indices, such as the clustering coefficient [35] (labeled as the C index), the extended neighborhood coreness centrality [15] (labeled as the C_{nc+} index), the weight degree centrality method (labeled as the W_{DC} index), the extended weight degree centrality method [12] (labeled as the EW_{DC} index), the extended gravity method [16] (labeled as the G_+ index), and the semi-local centrality method [11] (labeled as the SL index). Here, we briefly review the definitions of the six indices that will be discussed in this work in Table 2.

3.4. Experiment Results

When deliberately attacking 5% ~ 10% of important nodes in a complex network, the entire network may be paralyzed. Therefore, it is necessary to identify the top-ranking important nodes for protection. The importance evaluation index that can better distinguish the differences between nodes should have a higher resolution. We will use the monotonicity M of ranking list I to quantify the resolution of different indices, as follows [15]:

Table 2. List of contrast indices in the experiments

Index	Calculation process	Descriptions
Clustering Coefficient	$C(i) = 2E_i / (k_i \cdot (k_i - 1))$	1. k_i is the degree of node i .
C_{nc+}	$C_{nc}(i) = \sum_{j \in \Gamma(i)} ks(j)$, $C_{nc+}(i) = \sum_{j \in \Gamma(i)} C_{nc}(j)$	2. E_i is the number of edges actually existing between the k_i nodes of node i .
W_{DC}	$K_i = \sum_{j \in \Gamma(i)} k_j$, $W_{DC}(i) = (K - k_i) \cdot k_i^{ \Gamma(i) }$	3. $\Gamma(i)$ is the nearest neighborhood of node i .
EW_{DC}	$ED_i = \sum_{j \in \Gamma(i)} W_{DC}(j)$ $EW_{DC}(i) = (ED_i - W_{DC}(i)) \cdot W_{DC}(i)^{ j }$	4. $ks(i)$ is the k -shell value of node i .
G_+	$G(i) = \sum_{j \in \Gamma(i)} ks(i)ks(j) / d_{ij}^2$, $G_+(i) = \sum_{j \in \Gamma(i)} G(j)$	5. r is the assortativity coefficient of the network.
SL	$Q(s) = \sum_{j \in \Gamma(s)} N(j)$, $SL(i) = \sum_{s \in \Gamma(i)} Q(s)$	6. d_{ij} is the shortest path distance between node i and node j .
		7. $N(j)$ is the number of the nearest and the next nearest neighbors of node j .

$$M(I) = \left[1 - \frac{\sum_{i \in I} N_i (N_i - 1)}{N_p (N_p - 1)} \right]^2 \quad (8)$$

In Equation (8), N_i denotes the number of nodes with the same index value i , and N_p represents the number of nodes selecting a certain proportion p . If $M(I) = 1$, this means that the order of importance of the N_p nodes is completely monotonic, and each node is assigned a different index value. Otherwise, $M(I) = 0$ indicates that the index values of all nodes are the same, and the node importance cannot be completely distinguished. In Table 3, when p is approximately equal to 25%, the resolution M values of the seven ranking indices are calculated respectively. In the Karate network, $M(C) = 0$ indicates that the index values of the important nodes identified by the clustering coefficients are the same, the difference of the nodes cannot be distinguished, and the importance of the nodes cannot be accurately determined. It is not difficult to see from Table 3 that the M values of the KLD index, W_{DC} index and EW_{DC} index are all one in all four networks, indicating that the KLD index, W_{DC} index, and EW_{DC} index have higher resolution in identifying important nodes, and they can distinguish the differences among nodes very well.

Table 3. The monotonicity of these seven indices

Network	$M(KLD)$	$M(C)$	$M(C_{nc+})$	$M(W_{DC})$	$M(EW_{DC})$	$M(G_+)$	$M(SL)$
Karate network	1.0000	0	0.9556	1.0000	1.0000	1.0000	0.9778
Dolphin social network	1.0000	0.7845	0.9253	1.0000	1.0000	1.0000	0.9623
Books about US politics	1.0000	0.8183	0.9634	1.0000	1.0000	1.0000	0.9877
Neural network	1.0000	0.9424	1.0000	1.0000	1.0000	0.9985	0.9633

Mining a set of important nodes in a complex network plays a crucial role in understanding the structure and function of the network. In order to further verify the effect of the KLD method to identify important nodes, we first use the C index, C_{nc+} index, W_{DC} index, EW_{DC} index, G_+ index, SL index, and KLD index to calculate the node importance of four different networks. According to the ranking results of the seven indices, we remove the 25% top important nodes of each index in different networks and simulate the changes of the scale of the maximal connected subgraph and network efficiency when the networks are subjected to deliberate attacks, thereby evaluating the accuracy of each metric index.

In the experiment of simulating the influence of the deliberate attack network on the maximal connected coefficient, we use the C index, C_{nc+} index, W_{DC} index, EW_{DC} index, G_+ index, SL index, and KLD index to rank the importance of nodes in four real networks respectively and calculate the maximal connected coefficient after removing the top important nodes. The experimental results are shown in Table 4(a)-(d), and $\mu(n)$ represents the value of the maximal connected coefficient after the n top important nodes are attacked.

In Table 4(a)-(d), the decrease of the maximal connected coefficient in these four different networks caused by the KLD index is the most obvious. Especially in the Karate network from Table 4(a), the KLD index shows a better attack effect than other indices in the initial process of network attacks. That is, when the four top important nodes and ten top important nodes are deleted respectively, the corresponding maximal connected coefficients of the network are 29.41% and 8.82%, which are significantly lower than the maximal connected coefficients corresponding to other indices. In Table 4(b)-(d), the changes of the maximal connected coefficient corresponding to the KLD index, G_+ index, and EW_{DC} index are relatively synchronous, indicating that they can mine to the top ranked important nodes early and accurately. However, it is found that

the maximal connected coefficient of some networks does not decrease with the deletion of important nodes. For example, in the Karate network, when attacking the nodes with 11.76% ~ 23.53% importance ranking by the G_+ index, the maximal connected coefficient of the network is still 47.06%, which indicates that the nodes identified by the index are not very important. Among the four networks, the C index has the least efficiency in disrupting the network, that is, it has limited ability to distinguish the importance of network nodes.

Table 4(a). Values of the maximal connected coefficient after deliberately attacking the top important nodes identified by seven indices respectively in Karate network

Index	$\mu(2)$	$\mu(3)$	$\mu(4)$	$\mu(5)$	$\mu(6)$	$\mu(7)$	$\mu(8)$	$\mu(9)$	$\mu(10)$
C	94.12%	91.18%	88.24%	85.29%	82.35%	79.41%	76.47%	73.53%	70.59%
C_{nc+}	94.12%	73.53%	70.59%	67.65%	64.71%	58.82%	55.88%	52.94%	41.18%
W_{DC}	76.47%	70.59%	29.41%	23.53%	23.53%	23.53%	17.65%	17.65%	17.65%
EW_{DC}	76.47%	70.59%	29.41%	23.53%	23.53%	23.53%	23.53%	17.65%	17.65%
G_+	76.47%	70.59%	47.06%	47.06%	47.06%	47.06%	47.06%	23.53%	20.59%
SL	94.12%	91.18%	88.24%	85.29%	82.35%	79.41%	76.47%	73.53%	70.59%
KLD	76.47%	58.82%	29.41%	23.53%	17.65%	17.65%	14.71%	11.76%	8.82%

Table 4(b). Values of the maximal connected coefficient after deliberately attacking the top important nodes identified by seven indices respectively in Dolphin social network

Index	$\mu(1)$	$\mu(3)$	$\mu(5)$	$\mu(6)$	$\mu(8)$	$\mu(9)$	$\mu(11)$	$\mu(12)$	$\mu(14)$
C	98.39%	95.16%	91.94%	90.32%	87.10%	85.48%	77.42%	75.81%	72.58%
C_{nc+}	98.39%	95.16%	91.94%	90.32%	87.10%	85.48%	79.03%	74.19%	70.97%
W_{DC}	98.39%	95.16%	90.32%	88.71%	85.48%	83.87%	79.03%	77.42%	70.97%
EW_{DC}	98.39%	95.16%	87.10%	83.87%	80.65%	79.03%	74.19%	74.19%	67.74%
G_+	98.39%	95.16%	91.94%	88.71%	83.87%	82.26%	69.35%	66.13%	62.90%
SL	98.39%	95.16%	91.94%	90.32%	87.10%	85.48%	80.65%	79.03%	75.81%
KLD	95.16%	87.10%	82.26%	79.03%	74.19%	72.58%	69.35%	67.74%	61.29%

Table 4(c). Values of the maximal connected coefficient after deliberately attacking the top important nodes identified by seven indices respectively in Books about US politics network

Index	$\mu(2)$	$\mu(5)$	$\mu(8)$	$\mu(10)$	$\mu(13)$	$\mu(15)$	$\mu(18)$	$\mu(21)$	$\mu(24)$
C	98.10%	95.24%	92.38%	90.48%	87.62%	85.71%	82.86%	80.00%	77.14%
C_{nc+}	98.10%	95.24%	92.38%	90.48%	87.62%	85.71%	82.86%	80.00%	77.14%
W_{DC}	98.10%	95.24%	92.38%	90.48%	86.67%	84.76%	79.05%	75.24%	72.38%
EW_{DC}	98.10%	95.24%	92.38%	90.48%	86.67%	84.76%	79.05%	75.24%	58.10%
G_+	98.10%	95.24%	92.38%	90.48%	88.57%	83.81%	77.14%	73.33%	41.90%
SL	98.10%	95.24%	92.38%	90.48%	87.62%	85.71%	82.86%	79.05%	71.43%
KLD	98.10%	95.24%	92.38%	90.48%	87.62%	83.81%	78.10%	73.33%	56.19%

Table 4(d). Values of the maximal connected coefficient after deliberately attacking the top important nodes identified by seven indices respectively in Neural Network

Index	$\mu(14)$	$\mu(21)$	$\mu(29)$	$\mu(36)$	$\mu(44)$	$\mu(51)$	$\mu(59)$	$\mu(66)$	$\mu(74)$
C	95.29%	92.93%	90.24%	87.88%	85.19%	82.83%	80.13%	77.78%	75.08%
C_{nc+}	95.29%	92.93%	86.87%	84.51%	84.18%	82.49%	79.80%	77.10%	74.41%
W_{DC}	91.92%	89.56%	86.53%	84.18%	81.48%	79.12%	76.43%	74.07%	71.38%
EW_{DC}	91.92%	89.56%	86.53%	84.18%	81.48%	79.12%	76.43%	74.07%	68.69%
G_+	91.92%	89.23%	86.53%	84.18%	81.48%	79.12%	76.43%	73.06%	70.37%
SL	95.29%	92.93%	90.24%	87.88%	85.19%	79.12%	76.77%	76.77%	74.75%
KLD	89.56%	87.21%	83.50%	80.81%	77.10%	73.74%	70.71%	68.35%	64.98%

Figure 2(a)-(d) reflects the change in descending ratio of network efficiency e after removing a certain proportion of the top important nodes by using different metric indices. The worse the network connectivity after removing important nodes, the more obvious the downward trend of network efficiency. If the selected nodes have the same index values (such as C values), then N random experiments (N indicates a total number of nodes repeated) are carried out, and the average of network efficiency after running the results is taken.

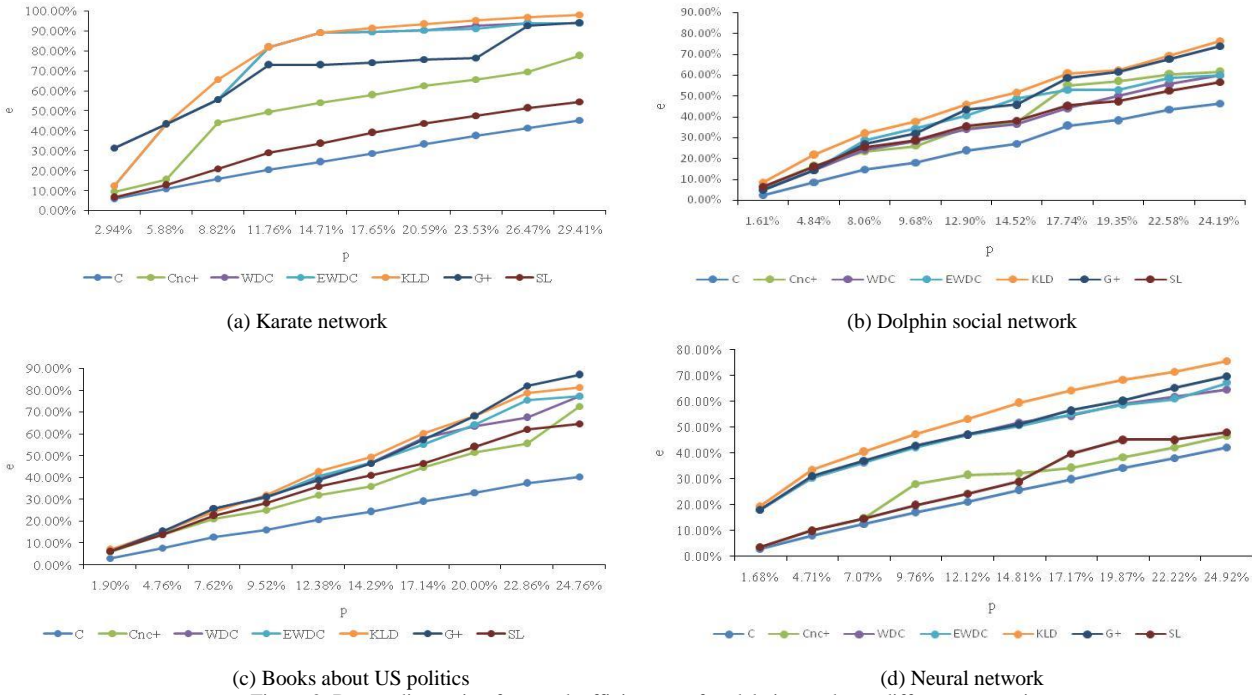


Figure 2. Descending ratio of network efficiency e after deleting nodes at different proportions p

In Figure 2(a), the W_{DC} index, EW_{DC} index, G_+ index, and KLD index can find the relatively important nodes in the network earlier and more accurately than the other three indices. The two curves of the EW_{DC} index and W_{DC} index basically coincide. When $p = 23.53\%$, the network descending ratio of the W_{DC} index and EW_{DC} index are 92.66% and 91.14%, respectively, and the two curves begin to separate from each other. This shows that the W_{DC} index finds the more important nodes earlier than the EW_{DC} index. The sudden increase in the slope of the curve corresponding to the index indicates that the index may identify the top important nodes. For example, when $p = 8.82\%$, the descending ratio of the network efficiency corresponding to the C_{nc+} index increases rapidly, indicating that the index identifies the important nodes. The SL index and C index perform relatively poorly. It is not difficult to find that in the Karate network, the ability of these indices to identify the important nodes from strong to weak are the KLD index, W_{DC} index, EW_{DC} index, G_+ index, C_{nc+} index, SL index, and C index.

In Figure 2(b), the experimental results show that the KLD index is significantly better than the other six indices. The overall synchronization of the KLD index and G_+ index is better. When $p = 24.19\%$, the network descending ratio of the KLD index and G_+ index is 76.48% and 73.97%, respectively. The performance of the EW_{DC} index in identifying important nodes is good in the early stage, but the slope of the curve decreases in the later stage, which indicates that there is some weakness in identifying important nodes. When $p = 17.74\%$, the network descending ratio of the EW_{DC} index and C_{nc+} index is 53.14% and 55.20%, respectively, and the C_{nc+} index identifies relatively important nodes earlier than the EW_{DC} index does. The SL index, W_{DC} index, and C index perform relatively poorly. It is not difficult to find that in the Dolphin social network, the ability of these indices to identify the important nodes from strong to weak are the KLD index, G_+ index, EW_{DC} index, C_{nc+} index, SL index, W_{DC} index, and C index.

In Figure 2(c), the KLD index and SL index can dig out important nodes earlier. When $p = 1.90\%$, the network descending ratio corresponding to the KLD index and SL index are 6.75% and 6.37%, respectively. However, the SL index does not perform well when identifying important nodes next. The curves corresponding to the W_{DC} index and EW_{DC} index are completely higher than the curve corresponding to the SL index. When $p = 4.76\%$, the network descending ratio of the SL index, W_{DC} index, and EW_{DC} index are 13.83%, 14.98%, and 14.61%, respectively. The C index has the worst performance in identifying important nodes. It is not difficult to see that in Books about US politics, the ability of these

indices to identify the important nodes from strong to weak are the KLD index, G_+ index, EW_{DC} index, W_{DC} index, SL index, C_{nc+} index, and C index.

It can be easily seen from Figure 2(d) that the curves corresponding to the SL index, C_{nc+} index, and C index are significantly lower than the curves corresponding to the other four indices, indicating that their ability to identify important nodes is poor. In the Neural network, the ability of these indices to identify the important nodes from strong to weak are the KLD index, G_+ index, W_{DC} index, EW_{DC} index, C_{nc+} index, SL index, and C index.

In Table 4 and Figure 2, for networks of different scales and different structural characteristics, the KLD index is more stable than the other six indices, and it can more accurately measure the node importance in complex networks. The reason for this result is that the KLD index comprehensively evaluates the node importance in the network from the perspective of the characteristics of the scale, tightness, and topology of the node and its neighbors. The other six indices are mainly based on a single attribute to evaluate the node importance, which is easily affected by the scale and topology of the network, resulting in experimental instability. From four different networks, we can also see that the KLD index, W_{DC} index, and WE_{DC} index are relatively stable in identifying important nodes, which are consistent with the M value calculated in Table 3. The more different the index values, the more accurate the ranking results. These seven indices have their own advantages and disadvantages in different networks, but the KLD index has the best overall effect, can effectively distinguish the most important nodes, and has strong versatility.

4. Conclusions

Identifying the important nodes of complex networks is helpful to improve the security protection capability of network hub nodes and enhance the network invulnerability and structural stability. Based on the degree attributes of nodes and their neighbors, this paper introduces two attributes of local network density and the assortativity coefficient, and it proposes a node importance evaluation method for multi-attribute evaluation. The method comprehensively considers the scale, tightness, and topology of nodes and their neighbors and makes up for the limitations of single attributes for evaluating the importance of network nodes. We use the KLD index and other six metrics to rank the node importance in four real networks and calculate the monotonicity index M value corresponding to each index. We find that the KLD index has higher resolution in distinguishing the differences between nodes. This paper chooses a certain proportion of nodes to carry out deliberate attack experiments on four real networks respectively. By comparing the experimental results of two evaluation indices of maximal connectivity coefficient and network efficiency, the KLD index performs better than the G_+ index, EW_{DC} index, W_{DC} index, C_{nc+} index, SL index, and C index, which verifies the validity and applicability of the KLD index. In this paper, the importance of nodes is analyzed from the structural point of view. In the future, we will focus on the measurement of node importance combined with dynamic characteristics and network structure.

Acknowledgements

We would like to thank the support from NSFC (Funding number 61402126, 61672179, and 61370083). Special thanks also go to the Youth Science Foundation (Number QC2016083) and Postdoctoral Support (Number LBH-Z14071) provided by the Heilongjiang Province. Finally, we thank the Doctoral Program of Higher Education of China for their support (Number 20122304110012).

References

1. D. J. Watts and S. H. Strogatz, "Collective Dynamics of Small-World Networks," *Nature*, Vol. 393, No. 6684, pp. 440-442, June 1998
2. A. L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, Vol. 286, No. 5439, pp. 509-512, October 1999
3. M. Canini, D. Venzano, P. Peres íni, D. Kostić, and J. Rexford, "A NICE Way to Test Open-Flow Applications," in *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, pp. 10, USENIX Association, Berkeley, April 2012
4. R. P. Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters*, Vol. 86, No. 14, pp. 3200-3203, April 2000
5. T. Rogers and A. J. McKane, "Modes of Competition and the Fitness of Evolved Populations," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, Vol. 92, No. 3, pp. 032708, September 2015
6. R. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling Cascading Failures in The North American Power Grid," *European Physical Journal B-Condensed Matter*, Vol. 46, No. 1, pp. 101-107, July 2005
7. G. Z. Wang, Y. J. Cao, Z. J. Bao, and Z. X. Han, "A Novel Local-World Evolving Network Model for Power Grid," *Acta Physica Sinica*, Vol. 58, No. 6, pp. 3597-3602, July 2009

8. T. Opsahl, F. Agneessens, and J. Skvoretz, "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths," *Social Networks*, Vol. 32, No. 3, pp. 245-251, July 2010
9. X. L. Ren and L. Y. Lü, "Review of Ranking Nodes in Complex Networks," *Chinese Science Bulletin*, Vol. 59, No. 13, pp. 1175, January 2014
10. Y. Moreno, M. Nekovee, and A. F. Pacheco, "Dynamics of Rumor Spreading in Complex Networks," *Physical Review E Statistical Nonlinear and Soft Matter Physics*, Vol. 69, No. 2, pp. 066130, June 2004
11. D. Chen, L. Lü, M. S. Shang, Y. C. Zhang, and T. Zhou, "Identifying Influential Nodes in Complex Networks," *Physica A Statistical Mechanics & its Applications*, Vol. 391, No. 4, pp. 1777-1787, February 2012
12. Y. Liu, B. Wei, Y. Du, F. Y. Xiao, and Y. Deng, "Identifying Influential Spreaders by Weight Degree Centrality in Complex Networks," *Chaos Solitons & Fractals*, Vol. 86, pp. 1-7, May 2016
13. M. Kitsak, L. K. Gallos, S. Havlin, and F. Lilijeros, "Identifying Influential Spreaders in Complex Networks," *Nature Physics*, Vol. 6, No. 11, pp. 888-893, January 2010
14. A. Zeng and C. J. Zhang, "Ranking Spreaders by Decomposing Complex Networks," *Physics Letters A*, Vol. 377, No. 14, pp. 1031-1035, June 2013
15. J. Bae and S. Kim, "Identifying and Ranking Influential Spreaders in Complex Networks by Neighborhood Coreness," *Physica A Statistical Mechanics & its Applications*, Vol. 395, No. 4, pp. 549-559, February 2014
16. L. L. Ma, C. Ma, H. F. Zhang, and B. H. Wang, "Identifying Influential Spreaders in Complex Networks based on Gravity Formula," *Physica A Statistical Mechanics & its Applications*, Vol. 451, pp. 205-212, June 2016
17. Y. Liu, M. Tang, T. Zhou, and Y. H. Do, "Core-Like Groups Result in Invalidation of Identifying Super-Spreader by K-Shell Decomposition," *Scientific Reports*, Vol. 5, pp. 9602, May 2015
18. Y. Liu, M. Tang, T. Zhou, and Y. H. Do, "Improving the Accuracy of the K-Shell Method by Removing Redundant Links: From a Perspective of Spreading Dynamics," *Scientific Reports*, Vol. 5, pp. 9602, August 2015
19. K. I. Goh, E. Oh, B. Kahng, and D. Kim, "Betweenness Centrality Correlation in Social Networks," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, Vol. 67, No. 1, pp. 017101, January 2003
20. G. Sabidussi, "The Centrality of a Graph," *Psychometrika*, Vol. 31, No. 4, pp. 581-603, December 1966
21. M. Rutter and G. W. Brown, "The Reliability and Validity of Measures of Family Life and Relationships in Families Containing a Psychiatric Patient," *Social Psychiatry*, Vol. 1, No. 1, pp. 38-53, August 1966
22. Z. M. Ren, F. Shao, J. G. Liu, and Q. Guo, "Node Importance Measurement based on the Degree and Clustering Coefficient Information," *Acta Physica Sinica*, Vol. 62, No. 12, pp. 505, June 2013
23. J. N. Yang, J. G. Liu, and G. Qiang, "Node Importance Identification for Temporal Network based on Inter-Layer Similarity," *Acta Physica Sinica*, Vol. 67, No. 4, pp. 048901, January 2018
24. C. Moore, G. Ghoshal, and M. E. Newman, "Exact Solutions for Models of Evolving Networks with Addition and Deletion of Nodes," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, Vol. 74, No. 2, pp. 036121, September 2006
25. S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks*, Vol. 56, No. 18, pp. 3825-3833, December 2012
26. L. Y. Lü, Y. C. Zhang, C. H. Yeung, and T. Zhou, "Leaders in Social Networks, the Delicious Case," *PLoS ONE*, Vol. 6, No. 6, pp. e21202, December 2011
27. Y. R. Gu and Z. Y. Zhu, "Node Ranking in Complex Networks based on LeaderRank and Modes Similarity," *Journal of University of Electronic Science & Technology of China*, Vol. 46, No. 2, pp. 441-448, March 2017
28. E. F. Codd, "A Relational Model for Large Shared Data Banks," *Communications of the ACM*, Vol. 13, pp. 377-387, January 1970
29. D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Sloaten, and S. M. Dawson, "The Bottleneck Dolphin Community of Doubtful Sound Features a Large Proportion of Long-Lasting Associations can Geographic Isolation Explain this Unique Trait?" *Behavioral Ecology and Sociobiology*, Vol. 54, No. 4, pp. 396-405, September 2003
30. J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The Structure of The Nervous System of the Nematode *Caenorhabditis Elegans*," *Philosophical Transactions of the Royal Society B Biological Sciences*, Vol. 314, No. 1165, pp. 1-340, November 1986
31. Y. C. Lai, A. E. Motter, and T. Nishikawa, "Attacks and Cascades in Complex Networks," *Complex Networks*, Vol. 650, No. 650, pp. 299-310, August 2004
32. S. Dereich and P. Mörters, "Random Networks with Sublinear Preferential Attachment: The Giant Component," *Annals of Probability*, Vol. 41, No. 1, pp. 329-384, July 2013
33. Y. R. Ruan, S. Y. Lao, J. D. Wang, L. Bai, and L. D. Chen, "Node Importance Measurement based on Neighborhood Similarity in Complex Network," *Acta Physica Sinica*, Vol. 66, No. 3, pp. 038902, December 2017
34. D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha, "Eigenvector-based Centrality Measures for Temporal Networks," *Physics*, Vol. 15, No. 1, pp. 537-574, July 2015
35. P. W. Holland and S. Leinhardt, "Transitivity in Structural Models of Small Groups," *Small Group Research*, Vol. 2, No. 2, pp. 49-66, May 1977

Hui Xu is studying for a doctorate in the College of Computer Science and Technology at Harbin Engineering University. She works in Heilongjiang University of Chinese Medicine. Her research interests are social computing, data mining technology, and software theory.

Jianpei Zhang is a professor in the College of Computer Science and Technology at Harbin Engineering University. He is the director of the Institute of Computer Software and Theory of Harbin Engineering University. He has long been engaged in database theory and application, data mining technology, software theory, and other aspects of teaching and research work.

Jing Yang is a professor in the College of Computer Science and Technology at Harbin Engineering University and an expert of the Harbin Informatization Committee. Her main research interests include database theory and application, data mining technology, knowledge database system, and software theory.

Lijun Lun is a professor in the College of Computer Science and Information Engineering at Harbin Normal University. He teaches courses on operating systems and software engineering, object-oriented software engineering, advanced software engineering, and new software technologies. His research interests are software testing and software metrics.