Information Sciences 512 (2020) 549–562

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Identification of influencers in complex networks by local information dimensionality

Tao Wen, Yong Deng*

Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China

ARTICLE INFO

Article history:
Received 18 June 2019
Revised 30 September 2019
Accepted 5 October 2019
Available online 5 October 2019

Keywords:
Complex network
Influential nodes
Local dimension
Shannon entropy

ABSTRACT

The identification of influential spreaders in complex networks is a popular topic in studies of network characteristics. Many centrality measures have been proposed to address this problem, but most have limitations. In this paper, a method for identifying influencers in complex networks via the local information dimensionality is proposed. The proposed method considers the local structural properties around the central node; therefore, the scale of locality only increases to half of the maximum value of the shortest distance from the central node. Thus, the proposed method considers the quasilocal information and reduces the computational complexity. The information (number of nodes) in boxes is described via the Shannon entropy, which is more reasonable. A node is more influential when its local information dimensionality is higher. In order to show the effectiveness of the proposed method, five existing centrality measures are used as comparison methods to rank influential nodes in six real-world complex networks. In addition, a susceptible–infected (SI) model and Kendall's tau coefficient are applied to show the correlation between different methods. Experiment results show the superiority of the proposed method.

1. Introduction

Complex networks are a popular topic that has attracted researchers' attention in many fields [14] because they can be used as a detailed model for many real-world complex systems such as brain networks [7], message networks [30,39], human lives [9], and social systems [25,27]. Many structural properties of complex networks are affected by some special nodes, e.g., the scale-free [18], self-similarity [22], and fractal [33] properties of complex networks [23]. In order to measure networks' properties effectively, many studies have been conducted to find these nodes with special properties, e.g., finding the most similar node [34], identifying influential nodes [26,43], and predicting potential links [36]. In particular, nodes with the ability to be highly influential in complex networks have gradually attracted researchers' attention because they have a greater influence on the networks' properties and structure than most other nodes, as demonstrated for predicting a time series by a visibility graph [11], predicting a link by similar nodes [2], detecting communities in social networks [37], measuring the network complexity [50], and dividing the network structure [41,42].

In general, each network has a specific node importance ranking, and different identification methods consider different structural properties of the network, which would give different ranking lists. Many centrality measures have been proposed to identify influential nodes, and they can be divided into three categories[13]: neighborhood-based, path-based, and

* Corresponding author.
E-mail address: dengentropy@uestc.edu.cn (Y. Deng).

https://doi.org/10.1016/j.ins.2019.10.003

iterative -refinement centralities. These centralities have many classical measures such as the degree centrality (DC) [16], betweenness centrality (BC) [16], closeness centrality (CC) [5], and eigenvector centrality (EC) [5]. In addition to new measures such as the H-index centrality [44], those in optimal percolation theory [3] and evidence theory [15], the technique for order preference by similarity to the ideal solution (TOPSIS) [48], and other measures [31,46,47]. These centrality measures have been applied in various fields such as game theory [32], human cooperation [19], evolutionary games [20], relevant website ranking [49], and node synchronization [4,38]. However, these classical centrality measures have limitations. For example, the DC only concentrates on local information and does not consider global information. The BC and CC focus on global information, but their high computational complexity limits their application to large-scale complex networks. The EC cannot be used in asymmetric networks, which reduces its application. Recently, some new centrality measures have been proposed. For instance, Zareie *et al.* ranked influential nodes using the entropy [45]. Deng *et al.* proposed a local dimension to identify vital nodes [21]. Makse *et al.* traced the real information flow in social networks to find influential spreaders [29].

The entropy is a useful tool for measuring the information of a complex network; hence, it has been widely used in many applications, e.g., evaluations of the vulnerability [35], presentation of the dimension [17,24], dilemma experiments [6], data fusion [8,28], entanglement measures [40], and evidence theory [10,12]. In addition, the structure of a complex network, such as the nodes and links, can be seen as probability sets. Therefore, the structural properties can be effectively explored by the entropy, which provides a new approach to address problems in the network, including the identification of important nodes.

In this paper, a new centrality measure is proposed to identify influential nodes on the basis of the local information dimensionality. The proposed method considers the information in boxes through the Shannon entropy, which is more reasonable than classical measures. In contrast to previous methods, the scale of locality of the proposed method increases from one to half of the maximum value of the shortest distance, which can consider the quasilocal information and reduce the computational complexity. Nodes with a higher local information dimensionality are more influential in the complex network, which is the same as classical measures. To show the effectiveness and reasonability of the proposed method, six real-world complex networks are considered, and five existing centrality measures are applied as comparison methods. Further, a susceptible–infected (SI) model and Kendall's tau coefficient [13] are used to show the superiority of the proposed method and the relationship between different methods.

The rest of this paper is organized as follows. Section 2 introduces some existing centrality measures and concepts about complex networks. The proposed local information dimensionality is discussed in Section 3. Some numerical experiments are presented in Section 4 to illustrate the effectiveness and reasonability of the proposed method. The conclusions are discussed in Section 5.

2. Preliminaries

2.1. Shortest distance between any two nodes in complex networks

In a given complex network $G(N, V)$, N is the set of nodes, and V is the set of edges. The adjacency matrix of the network can be obtained from the topological properties of network (the relationships between the nodes and the edges). Then, the shortest distance matrix can be obtained when the shortest distances between any two nodes are calculated by the Dijkstra algorithm. The adjacency matrix and shortest distance matrix are the known information of complex networks and are solved in advance to facilitate later application. The shortest distance ω_{ij} between node i and node j is defined as follows:

$$\omega_{ij} = \min(e_{ik_1} + e_{k_1k_2} + \cdots + e_{k_mj}) \tag{1}$$

where k_1, k_2, $\cdots$, k_m are the node IDs and $e_{k_1k_2}$ is the edge between two nodes. $e_{k_1k_2} = 1$ indicates that there is an edge between two nodes, and $e_{k_1k_2} = 0$ is the opposite. Thus, the shortest path length between node i and node j is denoted by ω_{ij}, and the maximum value of the shortest distance from node i is

$$\kappa_i = \max_{j\in N, j\neq i}(\omega_{ij}) \tag{2}$$

The maximum value of the shortest distance κ_i is the scale of locality around node i, and it is different for different nodes.

2.2. Centrality measures

Some existing measures are introduced in this section such as the BC, CC, DC, EC, and local dimension (LD).

Definition 2.1. Betweenness Centrality (BC) [16]. The BC of node i is denoted by $C_B(i)$ and defined as follows:

$$C_B(i) = \sum_{s,t\neq i} \frac{g_{st}(i)}{g_{st}} \tag{3}$$

where g_{st} is the number of shortest paths between node s and node t and $g_{st}(i)$ is the number of shortest paths between node s and node t that pass through node i.

Definition 2.2. Closeness Centrality (CC) [5]. The CC of node i is denoted by $C_C(i)$ and defined as follows:

$$C_C(i) = \left(\sum_{j=1}^{|N|} \omega_{ij} \right)^{-1} \tag{4}$$

where ω_{ij} is the shortest distance from node i to node j that can be obtained by Eq. (1) and $|N|$ is the number of nodes.

Definition 2.3. Degree Centrality (DC) [16]. The DC of node i is denoted by $C_D(i)$ and defined as follows:

$$C_D(i) = \sum_{j=1}^{|N|} e_{ij} \tag{5}$$

where e_{ij} is the edge between node i and j In fact, the DC means the number of edges connected with the selected node.

Definition 2.4. Eigenvector Centrality (EC) [5]. A is a similarity matrix whose size is $|N| \times |N|$. The EC x_i of node i is the ith entry in the normalized eigenvector that belongs to A, and it is defined as follows:

$$Ax = \lambda x, x_i = u \sum_{j=1}^{|N|} a_{ij} x_j \tag{6}$$

where λ is the largest eigenvalue of A, $u = 1/\lambda$, and x_i is the sum of the similarity scores of the nodes that are connected with node i.

The LD [21] of node i is introduced in Section 2.3.

2.3. Local dimension

To explore the local structural properties of complex networks, Silva *et al.* proposed the LD of complex networks. A power-law distribution has been proven to exist in theoretical networks with special properties such as small-world properties and many real-world networks. Because the topological scale from each central node is different, the LD changes with the selection of the central node. Pu *et al.* [21] modified the LD to identify the vital nodes in complex networks. For a radius r, it has been found that the number of nodes $N_i(r)$ whose shortest distance from the central node is less than r follows the power law

$$N_i(r) \sim r^{D_i} \tag{7}$$

It can be easily found that the LD D_i of node i can be obtained from the slope of a log–log plot, and it is expressed as follows:

$$D_i = \frac{d}{d \ln r} \ln N_i(r) \tag{8}$$

where d is the symbol of derivative. The radius r increases from one to the maximum value of the shortest distance κ_i from node i, and the derivative of Eq. (8) is expressed as follows because of the discrete properties [1] of complex networks:

$$D_i = \frac{r}{N_i(r)} \frac{d}{dr} N_i(r) \tag{9}$$

$$D_i = r \frac{n_i(r)}{N_i(r)} \tag{10}$$

where $n_i(r)$ is the number of nodes whose shortest distance from the central node equals r. When a central node is chosen, the scale of locality of the central node can be determined, and the LD of the central node can be obtained from the slope of a log–log plot ($\ln N_i(r)$ vs. $\ln r$). Lastly, the importance of a node can be determined by the order of the LD. In contrast to the previous methods, a node with a lower LD is more influential in the network.

3. Proposed method

Many centrality measures have been proposed to identify the influential nodes in complex networks. Different methods consider different structural information in the network and have their own advantages and limitations. Because most previous methods concentrate on the global or local structure, the quasilocal structure around the selected node cannot be effectively recognized. In this paper, a new method is proposed to identify vital nodes on the basis of the local information dimensionality (LID) of each node in a complex network. The proposed method considers the quasilocal information around each node and reduces the computational complexity. The practicality and effectiveness of the proposed method are demonstrated with experiments comparing some real-world complex networks in Section 4. A flowchart for obtaining the LID of one selected node is shown in Fig. 1.

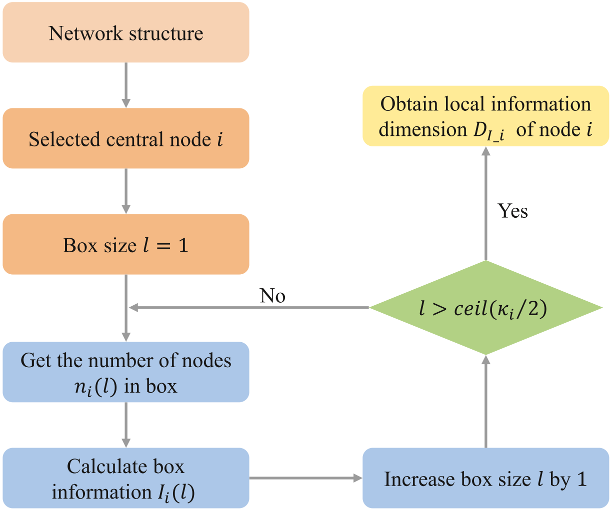


Fig. 1. Flowchart of the proposed method. The main step is to calculate the local information dimensionality D_{I_i} of one node from the structure of a complex network.

In this section, the LID of a complex network is proposed. The number of nodes in each box is considered using the Shannon entropy in the proposed method, which is more reasonable. Similar to the LD D_i, the LID D_{I_i} also considers the structural properties around node i in complex networks, and it is defined as follows:

$$D_{I\_i} = -\frac{d}{d\ln l} I_i(l) \tag{11}$$

where d is the symbol of derivative, l is the size of the box and $I_i(l)$ is the information in the box whose central node is node i with size l. In contrast to the classical LD, the information $I_i(l)$ in the selected box is considered using the Shannon entropy to describe the number of nodes in the box. In addition, the rule that governs the growth of the size of the box is different from the classical definition. The size of the box l grows from one to half of the maximum value of the shortest distance from the central node κ_i, i.e., $ceil(\kappa_i/2)$. The change in the size of the box means that the LID focuses on the quasilocal structure around the central node and reduces the computational complexity. The information $I_i(l)$ in each box is determined by the number of nodes in the box, and the number of selected nodes is considered using the Shannon entropy. Thus, the information in the box can indicate the node's properties more reasonably, and it is defined as follows:

$$I_i(l) = -p_i(l)\ln p_i(l) \tag{12}$$

where $p_i(l)$ is the probability that information is contained in a box whose central node is i for a given box size l, which is the ratio of the number of nodes in the box, $n_i(l)$, to the total number of nodes in the complex network, N, and can be obtained as follows:

$$p_i(l) = \frac{n_i(l)}{N} \tag{13}$$

Thus, the LID in Eq. (11) can be rewritten as follows:

$$D_{I\_i} = -\frac{d}{d\ln l}\left(-\frac{n_i(l)}{N}\ln\frac{n_i(l)}{N}\right) \tag{14}$$

From Eq. (14), the information in the box, $I_i(l)$, around the central node i is obtained from the number of nodes in the box by the Shannon entropy. The LID D_{I_i} of the selected node i is obtained from the slope of the line fitting the relationship between $I_i(l)$ and $\ln l$. Because of the network's discrete nature [1], the expression with the derivative in Eq. (14) can be rewritten as

$$\begin{aligned} D_{I\_i} &= -\frac{d}{d\ln l}(-p_i(l)\ln p_i(l)) \\ &= \frac{l}{1+\ln p_i(l)}\frac{d}{dl}p_i(l) \\ &\approx \frac{l}{1+\ln\frac{n_i(l)}{N}}\frac{N_i(l)}{N} \end{aligned} \tag{15}$$

where $N_i(l)$ is the number of nodes whose shortest distance from the central node i equals the box size l ($\omega_{ij} = l$) and $n_i(l)$ is the number of nodes whose shortest distance from the central node i is less than the box size l ($\omega_{ij} \le l$).

In the proposed method, the scale of locality $r_{\max}$ changes with the central nodes, which is defined as the half of the maximum value of the shortest distance from the central node κ_i, i.e., $r_{\max} = ceil(\kappa_i/2)$. The box size l increases from one to the scale of locality $r_{\max}$. The information in each box (the number of nodes in the box) is considered using the Shannon entropy. The LID of each node can be obtained from the slope of the box information $I_i(l)$ and the logarithm of the box size $\ln l$. Owing to the properties of the LID, the proposed method considers the information in the box more reasonably and reduces the computational complexity.

4. Experimental study

To show the effectiveness of the proposed method, six real-world complex networks and five comparison measures are used. These six complex networks are the USAir network, Jazz network, Karate network, Political blogs network, Facebook network, and (High Energy Physics - Theory) collaboration network from arXiv, which can be downloaded from http://vlado.fmf.uni-lj.si/pub/networks/data/ and http://snap.stanford.edu/data/. In addition, the collaboration network is chosen as the largest connected subgraph from the original network data. These five comparison measures (the BC, CC, DC, EC, and LD) were introduced in Section 2. The structural properties of these six networks are listed in Table 1. $|N|$ and $|V|$ are the numbers of nodes and edges respectively; $<k>$ and $k_{\max}$ are the average and maximum value of the degreeof centrality, respectively; and $<\omega>$ and $\omega_{\max}$ are the average and maximum value of the shortest distance in the network, respectively. Five experiments are implemented, including listing the top-10 node IDs to compare the differences in the top-10 node results obtained by different measures, the propagation based on the SI model to show the superior infection ability of the nodes obtained by the LID, the the relationship graph and Kendall's tau coefficient to show the similarity of the node rankings obtained by different measures and the SI model, and the running times of different measures to show the proposed method's low computational complexity.

4.1. Top-10 nodes

Firstly, the top-10 nodes in six real-world complex networks are identified by the LID and five other centrality measures, and the results are listed in Table 2. The nodes in color for the five existing measures are the same top-10 nodes identified by the LID. Because of the different consideration of information in the network, different centrality measures could give different lists of top-10 nodes. Thus, the numbers of same nodes for different measures can show the similarity of information considered by different methods, and the identification of more nodes by the LID, which also appear in the results of other measures, can increase its credibility. These unique nodes identified by the LID may indicate important changes in the propagation process. From Table 2, the same node with the most influence in the USAir network is obtained by the six different methods —node 118. The top-10 nodes obtained by the LID are the same as those obtained by the CC, and eight and seven of the same nodes are obtained between the LID and the DC and EC, respectively. The numbers of the same top-10 nodes obtained by the BC, LD, and LID are less than those of other methods, which are six. The results for the USAir network show that most of the top-10 nodes obtained by the other measures are obtained by the LID, which indicates its similarity and credibility.

For the Jazz network, the number of same top-10 nodes between the LID and the other measures is relatively small. There are only five same top-10 nodes among the CC, DC, EC, and LID; the BC and LD only have three and four of the same nodes, respectively, as the LID. The differences in the top-10 nodes between the LID and the other measures are large, but the unique nodes in the LID have a significant influence on the propagation process, which can show the importance of these nodes in complex networks. A detailed comparison is carried out with the experiments below.

The list of top-10 nodes of the Karate network in Table 2, is exactly the same using the LID and CC (this result is the same as the result for the USAir network). Many of the same top-10 nodes are observed for the DC, EC, and LID, which have the same nine nodes, thereby indicating the similarity between them. The numbers of same top-10 nodes between the LID and the BC and LD are seven and five, respectively.

Table 1
Topological properties of real-world networks.

Network	$\|N\|$	$\|V\|$	$<k>$	$k_{\max}$	$<\omega>$	$\omega_{\max}$
USAir	332	2126	12.8072	139	2.7381	6
Jazz	198	5484	27.6970	100	2.2350	6
Karate	34	78	4.5882	17	2.4082	5
Political blogs	1222	19021	27.3552	351	2.7375	8
Facebook	4039	88234	43.6910	1045	3.6925	8
Collaboration	8368	24827	5.7459	65	5.9454	18

Table 2
Top-10 nodes ranked by different centrality methods in six real-world complex networks. A node in color indicates that it also exists in the top-10 list obtained by the LID. The similarities in the top-10 nodes between different measures and the LID are provided.

Rank	USAir network						Jazz network					
	BC	CC	DC	EC	LD	LID	BC	CC	DC	EC	LD	LID
1	118	118	118	118	118	118	136	136	136	60	60	136
2	8	261	261	261	261	67	60	60	60	132	136	168
3	261	67	255	255	152	261	153	168	132	136	132	70
4	47	255	182	182	230	201	5	70	168	168	83	122
5	201	201	152	152	255	47	149	83	70	108	168	178
6	67	182	230	230	182	255	189	132	108	99	99	83
7	313	47	166	112	112	166	167	194	99	131	108	18
8	13	248	67	67	147	248	96	122	158	70	158	153
9	182	166	112	166	166	182	115	174	83	83	194	118
10	152	112	201	147	293	112	83	158	7	194	7	132

Rank	Karate network						Political blogs network					
	BC	CC	DC	EC	LD	LID	BC	CC	DC	EC	LD	LID
1	1	1	34	34	34	32	12	28	12	12	12	12
2	3	3	1	1	1	3	304	12	28	14	28	28
3	34	34	33	3	33	14	94	16	304	16	304	16
4	33	32	3	33	24	9	28	14	14	67	14	14
5	32	33	2	2	3	20	145	36	16	52	16	304
6	6	14	32	9	2	33	6	67	94	18	94	94
7	2	9	4	14	30	1	16	94	6	28	6	67
8	28	20	24	4	6	2	300	35	67	47	67	36
9	24	2	14	32	7	34	163	145	35	73	35	35
10	9	4	9	31	28	4	35	304	145	9	36	6

Rank	Facebook network						Collaboration network					
	BC	CC	DC	EC	LD	LID	BC	CC	DC	EC	LD	LID
1	107	107	107	801	107	107	3814	2448	187	4809	1763	2448
2	1684	58	1684	692	1684	1684	2448	3814	2448	2058	2448	6325
3	3437	428	1912	775	1912	1912	5489	7814	7935	814	5024	4772
4	1912	563	3437	749	4039	483	5380	7773	3814	6107	8313	6570
5	4039	1684	4039	841	3437	348	187	3097	2049	6219	7935	7773
6	58	171	2543	699	2543	414	1546	6301	7413	6010	4523	7814
7	1085	348	2347	788	2347	4039	1231	1499	2944	777	2049	7935
8	698	483	1888	743	2266	428	8301	7935	6095	6516	7413	3097
9	567	414	1800	750	1941	376	1805	4173	5489	5020	6017	2049
10	428	376	1663	802	1985	475	2944	5380	7773	4670	7773	1763

For the Political blogs network, the numbers of same top-10 nodes between the LID and the CC, DC, and LD are 9, 9, and 10 respectively, which are larger than the numbers of same top-10 nodes obtained by the BC (seven) and EC (five).

For the Facebook network, the largest number of same top-10 nodes as the LID is obtained for the CC (seven), which demonstrates the similarity of information considered by these two measures in this network. The results for the BC, DC, and LD are five, four, and four, respectively. However, there are no same top-10 nodes between the EC and the LID, which indicates that there is a large difference between these two measures.

Lastly, for the largest network —the collaboration network, the number of same top-10 nodes between the CC, LD, and LID is five, which is larger than those for other methods. The number of same top-10 nodes between the LID and the BC and EC are significantly different: one and zero, respectively.

Thus, these centrality measures consider the information in this network differently and give different lists of top-10 nodes. The DC has four of the same top-10 nodes as the LID, which is better than the BC and EC. From Table 2 and the discussion above, the CC and DC are two measures similar to the LID because they can obtain closer rankings than other methods. The BC and EC have different performance in these networks; some networks have similar rankings, and some are different. The reason why there are no same top-10 nodes between the EC and the LID is the influence of the network scale. The EC does not have good performance in large-scale networks because of its completely different list of top-10 nodes from other methods. In conclusion, the LID exhibits performance closer to existing measures for identifying top-10 nodes. A more detailed comparison with experiments is discussed below. Because the proposed method is modified from the LD and it focuses on a node's influence from different distances, the LD (the most related method) and CC (the same consideration factor) are chosen for comparison in the experiments below.

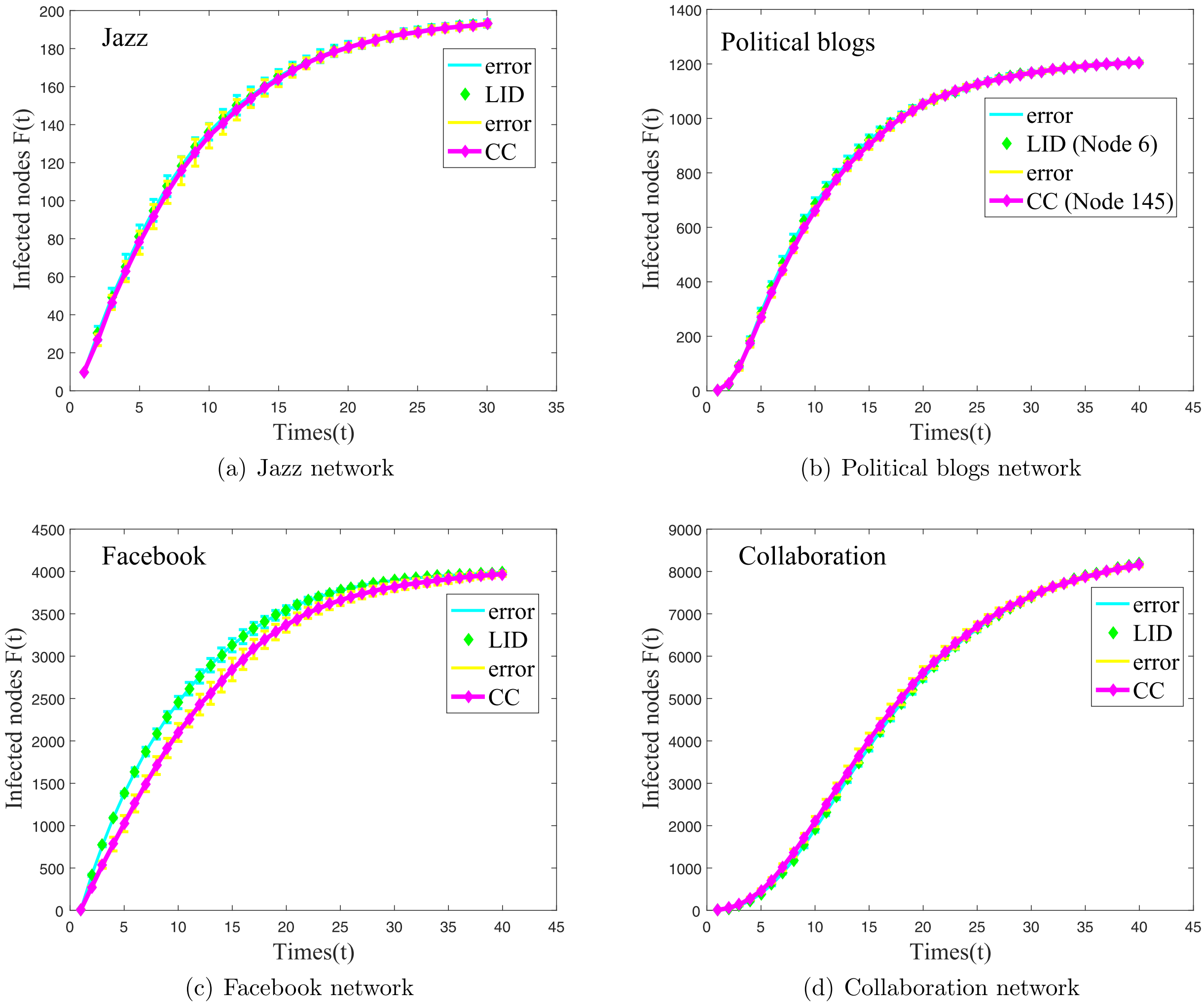


Fig. 2. Number of infected nodes for different initial nodes (top-10 nodes) obtained by the LID and CC in four networks. The infection ability of the top-10 nodes for the LID and CC is compared in this figure, and a higher number of infected nodes $F(t)$ indicates that the initial nodes have a higher infection ability. The results are obtained from 50 independent experiments with $\beta = 3$.

4.2. SI model

Detailed experiments were carried out to show which measure is more effective and reasonable. In the SI method, each node has two states: infected and susceptible. Some nodes were initially selected as infected nodes. A susceptible node becomes infected by these, and it can no longer return to the susceptible state. Additionally, the state of each node can only be affected by its neighboring nodes with a probability λ. In this study, the SI model is applied to measure the infection ability of some selected initial nodes, which is positively correlated with the degree of importance of a node. The top-10 nodes obtained by different methods are used as the initial infected nodes, and the rest of nodes are defined as susceptible nodes. Every time t, infected nodes have a spread rate $\lambda = (1/2)^{\beta}$ for infecting their neighboring susceptible nodes, and the total numbers of infected and susceptible nodes are equal to the number of nodes $|N|$ in complex networks. β has different settings for the different scales of the networks. After infection at time t, the initial nodes with a higher infection ability cause a greater number of infected nodes in the network, which can indicate the importance of these nodes. The number of infected nodes $F(t)$ at some specific time t is chosen as an indicator to measure the infection ability of the initial infected nodes. A higher number of infected nodes indicates that the infection ability of the initial nodes is stronger and that the initial nodes are more important.

The LD and CC are selected as comparison methods because of their consideration of information. The LID is compared with the CC and LD by the SI model described as follows. First, the initial infected nodes are chosen as the top-10 nodes obtained by different methods and are listed in Table 2. Then, the infection process lasts for a time t, and the number of infected nodes $F(t)$ is recorded. Lastly, every experiment is carried out independently and is repeated 50 times with $\beta = 3$. The results are the average of 50 experiments, which are shown in Fig. 2 and Fig. 3.

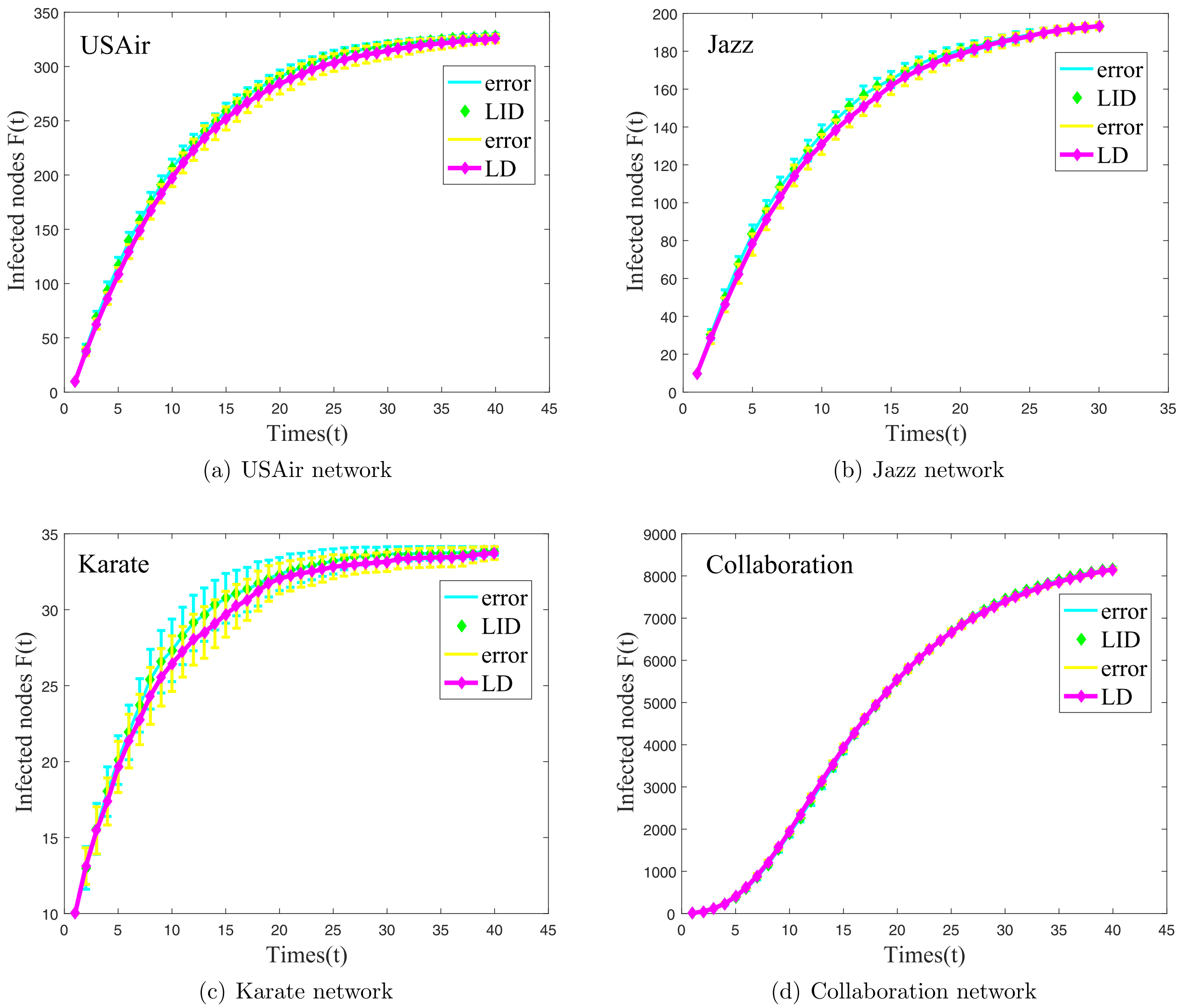
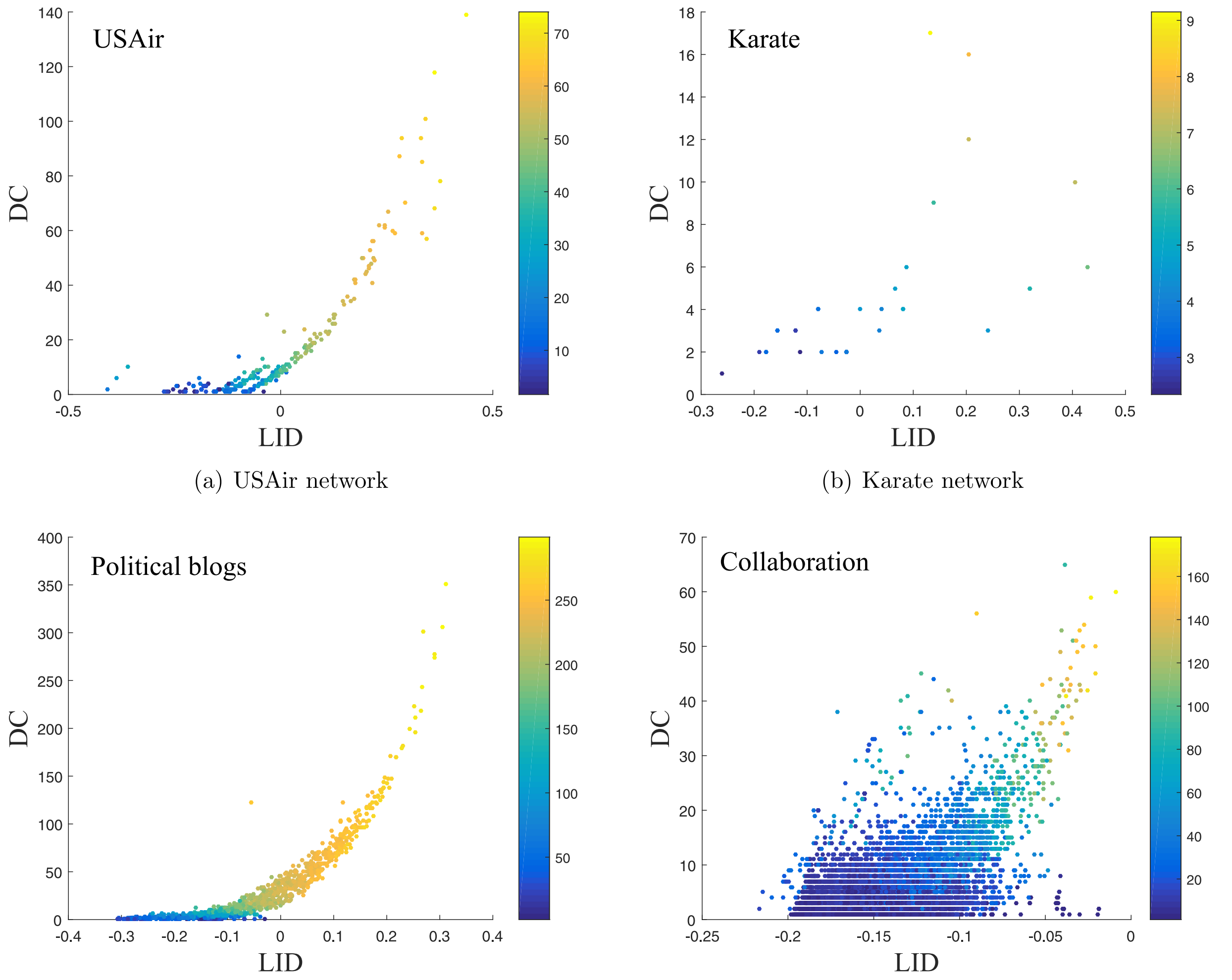


Fig. 3. Number of infected nodes for different initial nodes (top-10 nodes) obtained by the LID and LD in four networks. The infection ability of the top-10 nodes for the LID and LD is compared in this figure, and a higher number of infected nodes $F(t)$ indicates that the initial nodes have a higher infection ability. The results are obtained from 50 independent experiments with $\beta = 3$.

From Fig. 2, the number of infected nodes $F(t)$ increases with the transmission time and eventually reaches a stable value. Because the lists of top-10 nodes obtained by the CC for the USAir and Karate networks are the same as that for the LID, the other four networks are used in SI model to compare the LID and CC. In the Jazz network, the LID is slightly better than the CC, which can be seen for $t = 5$–20. In the Political blogs network, because the CC has nine of the same top-10 nodes as the LID, the one different node is chosen as the initial infected node to simulate the SI model to show the performance difference between the LID and the CC. The performance of the LID is better than that of the CC, as observed from the early time period in the SI model. In the Facebook network, the LID is clearly superior to the CC. $F(t)$ for the LID is larger than that for the CC over the entire transmission process, and the LID reaches a stable value earlier than the CC. In the collaboration network, the LID has a slightly lower $F(t)$ in the early time period, but it keeps up with the growth of the CC after $t = 30$. The results comparing the LD and LID are shown in Fig. 3. In the USAir network, the LID has a stronger spreading ability than the LD because the number of infected nodes obtained by the LID is larger than that of the LD in the middle time period. Moreover, the LID is clearly superior to the LD over the entire progression of the Jazz network. In the Karate network, the LID is obviously better than the LD for the average number of infected nodes and significantly more stable. In the collaboration network, the LD has a similar effectiveness as the LID because their curves almost overlap. Overall, the superiority of the LID is obvious in most of the SI experiments from the observations of $F(t)$ in the comparison between the CC, LD, and LID for different networks. In some cases, the LID has similar performance as that of other existing methods, which only have a slight advantage.



(a) USAir network (b) Karate network

(c) Political blogs network (d) Collaboration network

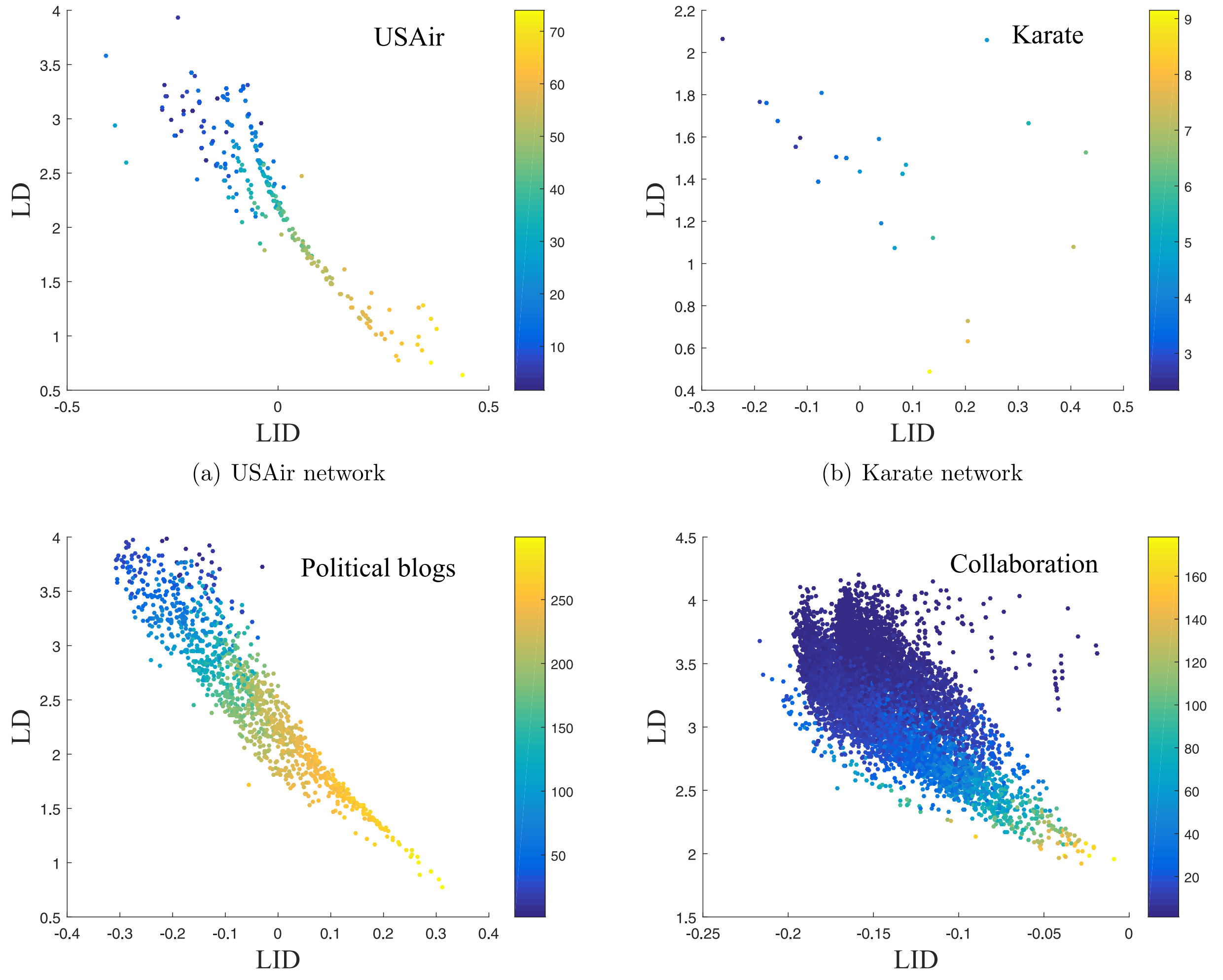
Fig. 4. Relationship between the LID and the DC for a spreading rate $\lambda = 0.05$in four networks. Each point represents one node in the network, and the color of a point represents the number of nodes $F(t)$ infected with the selected initial node at $t = 10$, which is obtained from 100 independent experiments. The color and the change in the value of the points show the correlation between the DC, the LID, and the SI model, and the monotonic changes show the similarity between these measures in the general trend.

4.3. Relationships between different measures

To find the relationship between the values obtained by different measures, a relationship graph between different methods is presented. The DC and LD are chosen for comparison. In the relationship graph, each point represents one node in a complex network. The values along the horizontal and vertical axes represent the LID value of each node and the DC or LD value of each node, respectively. The color of a point represents the infection ability of the selected node over 10 time steps ($F(10)$) when $\lambda = 0.05$ in the SI model (obtained by 100 independent experiments). When a node has a large LID and large DC or LD, these two methods have a positive correlation; a negative correlation is obtained when a node has a large LID and a small DC or LD. The detailed results are shown in Figs. 4 and 5.

From the correlation between the DC and the LID in Fig. 4, the nodes with large LIDs have large DCs, which means that the DC is positively correlated with the LID. Owing to the properties of the DC, there are many nodes with same degree of centrality, which can be seen in the figure. Thus, the LID has an obvious change, but the change in the DC is relatively small in the early time period, which demonstrates that the important nodes cannot be effectively identified with the small DC Moreover, there are many nodes with a small degree of centrality, which follows the scale-free feature of complex networks. Therefore, this phenomenon shows the superiority of the LID.

From Fig. 5, the relationship between the LID and the LD has a negative correlation; that is, nodes with a high $F(10)$ have a high LID but a low LD. This is because of the features [21] of the LD (a node with greater importance has a smaller LD).



(a) USAir network
(b) Karate network
(c) Political blogs network
(d) Collaboration network

Fig. 5. Relationship between the LID and the LD for a spreading rate $\lambda = 0.05$**in four networks.** Each point represents one node in the network, and the color of a point represents the number of nodes $F(t)$ infected with the selected initial node at $t = 10$, which is obtained from 100 independent experiments. The color and the change in the value of the points show the correlation between the LD, the LID, and the SI model, and the monotonic changes show the similarity between these measures in the general trend.

The relation between the LID and the LD is similar to a linear relation, which indicates that they would give similar rank lists.

In conclusion, the proposed method is the same as classical measures; a larger value for a measure indicates a stronger infection ability. The LID can achieve stable correlative performance with other centrality measures in different real-world complex networks. In addition, the LID can more effectively identify the nodes' importance with a small degree of centrality.

4.4. Kendall's tau coefficient

Kendall's tau coefficient [13] has been applied to measure the correlation between the centrality measures and the infection ability measured by the SI model [13]. Kendall's tau coefficient can measure the correlation between two different variables, and a higher Kendall's tau coefficient indicates that these two variables are more similar, which can obtain a more effective result.

The definition of Kendall's tau coefficient is as follows. For two random variables A and B, their ith combination is denoted by (A_i, B_i). When $A_i > A_j$ and $B_i > B_j$ or $A_i < A_j$ and $B_i < B_j$ simultaneously occur, (A_i, B_i) and (A_j, B_j) are considered concordant. (A_i, B_i) and (A_j, B_j) are considered discordant when $A_i > A_j$ and $B_i < B_j$ or $A_i < A_j$ and $B_i > B_j$ simultaneously occur. In addition, when $A_i = A_j$ and $B_i = B_j$, (A_i, B_i) and (A_j, B_j) are considered neither discordant nor concordant. Therefore, Kendall's tau coefficient τ is defined as follows:

$$\tau = \frac{n_c - n_d}{0.5n(n-1)} \tag{16}$$

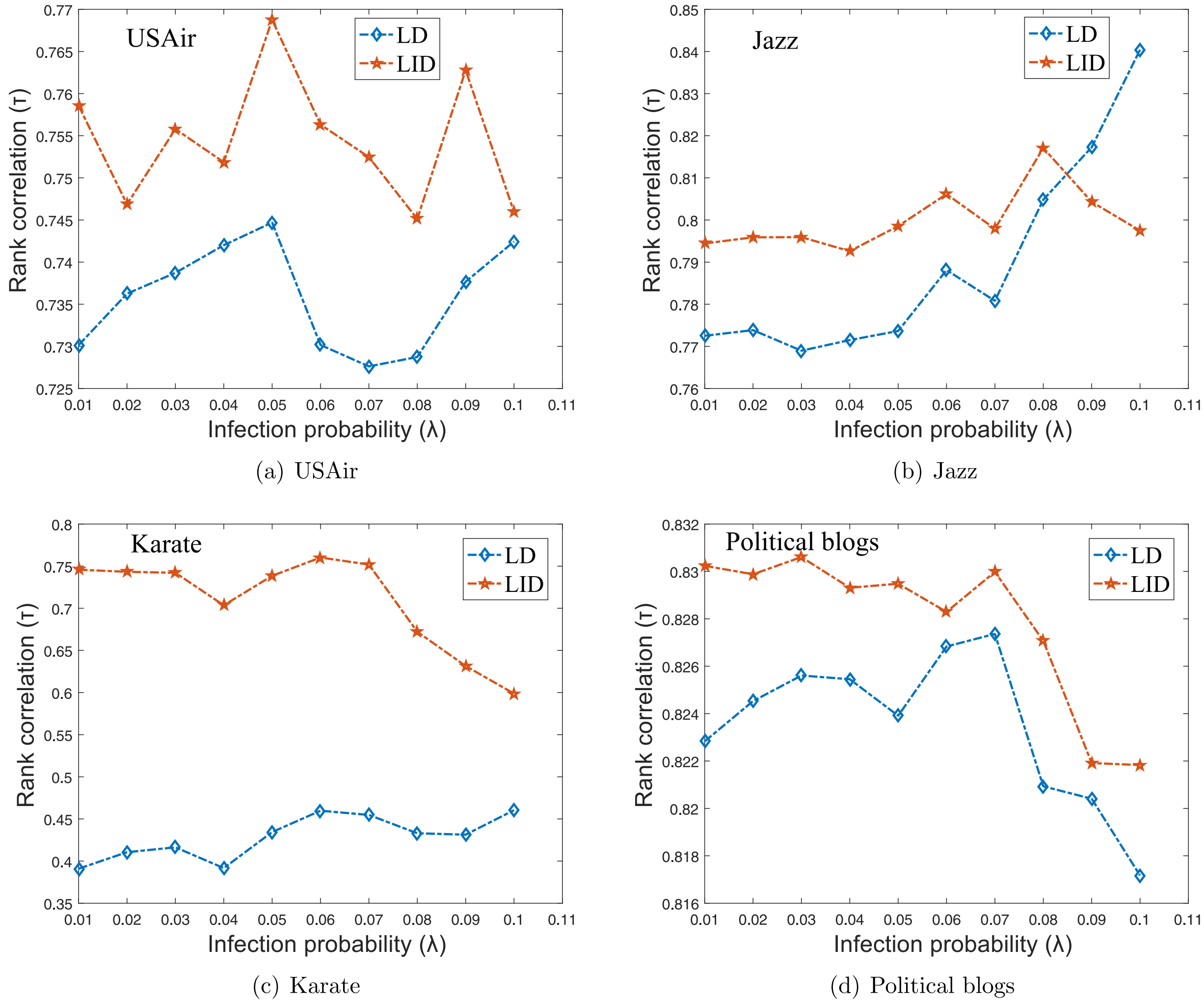


Fig. 6. Kendall's tau coefficient τ between the infection ability obtained by the SI model and the LID and LD for four networks. The infection ability of each node is obtained by the SI model from 100 independent experiments. A higher τ means that a method is more similar to the SI model given a change in the infection probability λ.

where n_c and n_d are the numbers of concordant and discordant combinations, respectively, and n is the number of combinations in the sequence. $\tau = 1$ indicates that the list ranking the nodes' importance obtained by different methods is the same as the list ranking the infection ability obtained by the SI model; $\tau = 0$ is the opposite case.

Kendall's tau coefficient τ between the LID and LD and the infection ability is compared. The infection ability of each node is represented by the number of infected nodes in 10 steps ($F(10)$) of the SI model. Additionally, different cases are considered in this experiment. The spreading rate λ in the SI model is varied from 0.01 to 0.1 to examine τ. The infection process is independently repeated 100 times, and τ is obtained by averaging. A larger τ indicates that the relationship between the infection ability and the centrality measure is more relevant

The results for Kendall's tau coefficient τ for four real-world complex networks are shown in Fig. 6. For the USAir network, τ does not have an obvious relationship with the change in λ, and the difference between τ is relatively small. In addition, τ for the LID is always larger than τ for the LD, which indicates the superiority of the proposed method. For the Jazz network, τ for the LID is larger than τ of the LD when λ increases from 0.01 to 0.08; then, τ of the LD rapidly increases after $\lambda = 0.08$, indicating that τ of the LD is larger than τ of the LID when $\lambda = 0.09, 0.1$. Hence, the performance of the proposed method is better most of the time. The values of τ for the Karate and Political blogs networks exhibit a downward trend. For the Karate network, τ of the LID is much larger than τ of the LD, indicating that the proposed method outperforms the LD. For the Political blogs network, the difference in the values of τ for the LID and LD is small, but τ of the LID is always larger than τ of the LD, which shows the superiority of the LID. In conclusion, τ between the LID and the infection ability is larger than τ between the LD and the infection ability in most cases. This means that the results obtained by the proposed method are more relevant to the classical infection ability, and the LID can maintain relatively stable correlative

Table 3
Running times (in seconds) of different centrality measures for different real-world networks.

Network	Time(BC)	Time(CC)	Time(DC)	Time(EC)	Time(LD)	Time(LID)
USAir	31.7043	0.0033	0.0007	0.0171	0.0201	0.0122
Jazz	11.1378	0.0016	0.0006	0.0086	0.0199	0.0058
Karate	0.1748	0.0004	0.0003	0.0009	0.0017	0.0006
Political blogs	611.2053	0.0435	0.0056	0.4676	0.2551	0.1252
Facebook	18635.5697	0.4985	0.0478	14.3498	2.9104	1.2109
Collaboration	291827.8567	4.3902	0.2290	193.3321	23.7418	11.5581

performance than the LD in most real-world complex networks. Thus, the proposed method is more effective for identifying influencers from this perspective.

4.5. Time consumption

The time consumption of different measures for different networks is presented. All centrality measures were calculated using MATLAB 2016a on a personal computer (PC) equipped with an Intel Core i7-5500U central processing unit (CPU) operating at 2.40 GHz and 8GB of random access memory (RAM). The method with a lower running time has a lower computational complexity. The running times of these measures are listed in Table 3. Based on the results, the DC has the lowest computational complexity on large- or small-scale networks. In contrast, the BC has the highest running time and far exceeds those of the other methods. The running time of the LID is approximately half of that of the LD and less than half in some cases. This is because the LID considers the information of nodes (quasilocal information) whose distance from the central node is less than half of the maximum value of the shortest distance, but the LD considers all of the nodes in the network. The running time of the EC is small for small-scale networks, but it rapidly increases for large-scale networks (as seen from the comparison with the LD). Additionally, the running time of the EC is 2–10 times that of the LID, which means that the LID has a relatively low computational complexity. The reason why the running time of the CC is smaller than that of the LID is that the CC only needs to sum the shortest distance from the central node to other nodes. In conclusion, the LID has a lower running time than most other methods, implying that the LID reduces the computational complexity.

5. Conclusion

In this paper, the influencers in complex networks are identified by the LID. The size of the box covering the central node grows from one to $ceil(d_i/2)$, and the number of nodes within the box is considered using the Shannon entropy, which can measure the information in the box. Then, the LID of the central node can be obtained by the correlation between the box information and the size of box. Finally, the ability of nodes to influence others can be ordered according to the value of the LID. Because of the rule governing the increase in the size of the box, the proposed method considers the quasilocal information around the central node and reduces the computational complexity. Compared with existing measures for real-world networks, the proposed method is more effective and reasonable, and experimental results demonstrate its superiority.

However, it can be improved. One inevitable problem is how to identify the ability of two spreaders to influence others when they have equal LIDs D_{I_i} (or other measures' results). In future research, the information in the box can be considered to be multiscale, which can achieve adequate consideration of information, and a better result can be obtained. Therefore, the framework of the dimension-based approach would be significantly improved for identifying influencers in complex networks.

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

The authors greatly appreciate the reviewers' suggestions and editor's encouragement. This work is partially supported by the National Natural Science Foundation of China (grant nos. 61973332, 61573290, and 61503237).

References

[1] E. Ben-Naim, H. Frauenfelder, Z. Toroczkai, Complex Networks, Springer Berlin Heidelberg, 2004.
[2] Z. Bu, Y.Y. Wang, H.J. Li, J.C. Jiang, Z. Wu, J. Cao, Link prediction in temporal networks: integrating survival analysis and game theory, Inf. Sci. 498 (2019) 41–61, doi:10.1016/j.ins.2019.05.050.
[3] G. Del Ferraro, A. Moreno, B. Min, F. Morone, U. Perez-Ramirez, L. Perez-Cervera, L.C. Parra, A. Holodny, S. Canals, H.A. Makse, Finding influential nodes for integration in brain networks using optimal percolation theory, Nat. Commun. 9 (2018) 12, doi:10.1038/s41467-018-04718.

[4] S.S. Feng, L. Wang, S.W. Sun, C.Y. Xia, Synchronization properties of interconnected network based on the vital node, Nonlinear Dyn. 93 (2) (2018) 335–347, doi:10.1007/s11071-018-4195-0.
[5] L.C. Freeman, Centrality in social networks conceptual clarification, Social Netw. 1 (3) (1979) 215–239, doi:10.1016/0378-8733(78)90021-7.
[6] L.K. Gallos, P. Barttfeld, S. Havlin, M. Sigman, H.A. Makse, Collective behavior in the spatial spreading of obesity, Scientif. Rep. 2 (2012a) 9, doi:10.1038/srep00454.
[7] L.K. Gallos, H.A. Makse, M. Sigman, A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks, Proc. Natl. Acad. Sci. USA 109 (8) (2012b) 2825–2830, doi:10.1073/pnas.1106612109.
[8] X.Z. Gao, Y. Deng, The generalization negation of probability distribution and its application in target recognition based on sensor fusion, Int. J. Distrib. Sensor Netw. 15 (5) (2019) 8, doi:10.1177/1550147719849381.
[9] D. Helbing, D. Brockmann, T. Chadefaux, K. Donnay, U. Blanke, O. Woolley-Meza, M. Moussaid, A. Johansson, J. Krause, S. Schutte, M. Perc, Saving human lives: what complexity science and information systems can contribute, J. Stat. Phys. 158 (3) (2015) 735–781, doi:10.1007/s10955-014-1024-9.
[10] W. Jiang, Y. Cao, X.Y. Deng, A novel Z-network model based on bayesian network and Z-number, IEEE Trans. Fuzzy Syst. (2019), doi:10.1109/TFUZZ.2019.2918999.
[11] F. Liu, Y. Deng, A fast algorithm for network forecasting time series, IEEE Access 7 (1) (2019) 102554–102560.
[12] Y.N. Liu, W. Jiang, A new distance measure of interval-valued intuitionistic fuzzy sets and its application in decision making, Soft Comput. 23 (2019) 10.1007/s00500–019–04332–5.
[13] L.Y. Lu, D.B. Chen, X.L. Ren, Q.M. Zhang, Y.C. Zhang, T. Zhou, Vital nodes identification in complex networks, Phys. Rep.-Rev. Sect. Phys. Lett. 650 (2016) 1–63, doi:10.1016/j.physrep.2016.06.007.
[14] S. Majhi, B.K. Bera, D. Ghosh, M. Perc, Chimera states in neuronal networks: a review, Phys. Life Rev. 28 (2019) 100–121, doi:10.1016/j.plrev.2018.09.003.
[15] H.M. Mo, Y. Deng, Identifying node importance based on evidence theory in complex networks, Physica A-Stat. Mech. Appl. 529 (2019) 11, doi:10.1016/j.physa.2019.121538.
[16] M.E.J. Newman, The structure and function of complex networks, Siam Rev. 45 (2) (2003) 167–256, doi:10.1137/s003614450342480.
[17] W. Pedrycz, A. Bargiela, Fuzzy fractal dimensions and fuzzy modeling, Inf. Sci. 153 (2003) 199–216, doi:10.1016/s0020-0255(03)00075-6.
[18] M. Perc, The matthew effect in empirical data, J. R. Soc. Interf. 11 (98) (2014) 15, doi:10.1098/rsif.2014.0378.
[19] M. Perc, J.J. Jordan, D.G. Rand, Z. Wang, S. Boccaletti, A. Szolnoki, Statistical physics of human cooperation, Phys. Rep.-Rev. Sect. Phys. Lett. 687 (2017) 1–51, doi:10.1016/j.physrep.2017.05.004.
[20] M. Perc, A. Szolnoki, Coevolutionary games-a mini review, Biosystems 99 (2) (2010) 109–125, doi:10.1016/j.biosystems.2009.10.003.
[21] J. Pu, X.W. Chen, D.J. Wei, Q. Liu, Y. Deng, Identifying influential nodes based on local dimension, Epl 107 (1) (2014) 6, doi:10.1209/0295-5075/107/10010.
[22] E. Rosenberg, Maximal entropy coverings and the information dimension of a complex network, Phys. Lett. A 381 (6) (2017a) 574–580, doi:10.1016/j.physleta.2016.12.015.
[23] E. Rosenberg, Minimal partition coverings and generalized dimensions of a complex network, Phys. Lett. A 381 (19) (2017b) 1659–1664, doi:10.1016/j.physleta.2017.03.004.
[24] E. Rosenberg, Non-monotonicity of the generalized dimensions of a complex network, Phys. Lett. A 381 (28) (2017c) 2222–2229, doi:10.1016/j.physleta.2017.05.014.
[25] A. Sheikhahmadi, M.A. Nematbakhsh, Identification of multi-spreader users in social networks for viral marketing, J. Inf. Sci. 43 (3) (2017) 412–423, doi:10.1177/0165551516644171.
[26] A. Sheikhahmadi, M.A. Nematbakhsh, A. Shokrollahi, Improving detection of influential nodes in complex networks, Physica A-Stat. Mech. Appl. 436 (2015) 833–845, doi:10.1016/j.physa.2015.04.035.
[27] A. Sheikhahmadi, M.A. Nematbakhsh, A. Zareie, Identification of influential users by neighbors in online social networks, Physica A-Stat. Mech. Appl. 486 (2017) 517–534, doi:10.1016/j.physa.2017.05.098.
[28] Y.T. Song, Y. Deng, A new method to measure the divergence in evidential sensor data fusion, Int. J. Distrib. Sensor Netw. 15 (4) (2019) 8, doi:10.1177/1550147719841295.
[29] X. Teng, S. Pei, F. Morone, H.A. Makse, Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks, Scientif. Rep. 6 (2016) 11, doi:10.1038/srep36043.
[30] X.Y. Wang, T.F. Zhao, Model for multi-messages spreading over complex networks considering the relationship between messages, Commun. Nonlinear Sci. Numer. Simul. 48 (2017) 63–69, doi:10.1016/j.cnsns.2016.12.019.
[31] Y.C. Wang, S.S. Wang, Y. Deng, A modified efficiency centrality to identify influential nodes in weighted networks, Pramana-J. Phys. 92 (4) (2019) 11, doi:10.1007/s12043-019-1727-1.
[32] Z. Wang, C.T. Bauch, S. Bhattacharyya, A. d'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathe, D.W. Zhao, Statistical physics of vaccination, Physics Reports-Review Section of Phys. Lett. 664 (2016) 1–113, doi:10.1016/j.physrep.2016.10.006.
[33] B. Wei, Y. Deng, A cluster-growing dimension of complex networks: from the view of node closeness centrality, Physica A-Stat. Mech. Appl. 522 (2019) 80–87, doi:10.1016/j.physa.2019.01.125.
[34] T. Wen, S.Y. Duan, W. Jiang, Node similarity measuring in complex networks with relative entropy, Commun. Nonlinear Sci. Numer. Simul. 78 (2019) 104867, doi:10.1016/j.cnsns.2019.104867.
[35] T. Wen, M.X. Song, W. Jiang, Evaluating topological vulnerability based on fuzzy fractal dimension, Int. J. Fuzzy Syst. 20 (6) (2018) 1956–1967, doi:10.1007/s40815-018-0457-8.
[36] Y. Xie, M.G. Gong, S.F. Wang, W.F. Liu, B. Yu, Sim2vec: node similarity preserving network embedding, Inf. Sci. 495 (2019a) 37–51, doi:10.1016/j.ins.2019.05.001.
[37] Y. Xie, X.M. Wang, D. Jiang, R.B. Xu, High-performance community detection in social networks using a deep transitive autoencoder, Inf. Sci. 493 (2019b) 75–90, doi:10.1016/j.ins.2019.04.018.
[38] W. Xing, P. Shi, H.Y. Song, Y.X. Zhao, L.Y. Li, Global pinning synchronization of stochastic delayed complex networks, Inf. Sci. 490 (2019) 113–125, doi:10.1016/j.ins.2019.03.042.
[39] D.G. Xu, X.Y. Xu, Y.F. Xie, C.H. Yang, Optimal control of an sivrs epidemic spreading model with virus variation based on complex networks, Commun. Nonlinear Sci. Numer. Simul. 48 (2017) 200–210, doi:10.1016/j.cnsns.2016.12.025.
[40] H.C. Yang, Y. Deng, J. Jones, Network division method based on cellular growth and physarum-inspired network adaptation, Int. J. Unconven. Comput. 13 (6) (2018) 477–491.
[41] H.C. Yang, Y. Deng, R. Mayne, A bio-inspired network design method for intelligent transportation, Int. J. Unconven. Comput. 14 (3) (2019a) 197–208.
[42] H.C. Yang, Y.J. liu, Q. Wan, Y. Deng, A bio-inspired optimal network division method, Physica A: Stat. Mech. Appl. 527 (2019b) 121259, doi:10.1016/j.physa.2019.121259.
[43] A. Zareie, A. Sheikhahmadi, A hierarchical approach for influential node ranking in complex social networks, Expert Syst. Appl. 93 (2018) 200–211, doi:10.1016/j.eswa.2017.10.018.
[44] A. Zareie, A. Sheikhahmadi, EHC: Extended h-index centrality measure for identification of users' spreading influence in complex networks, Physica A-Stat. Mech. Appl. 514 (2019) 141–155, doi:10.1016/j.physa.2018.09.064.
[45] A. Zareie, A. Sheikhahmadi, A. Fatemi, Influential nodes ranking in complex networks: an entropy-based approach, Chaos Solitons Fractals 104 (2017) 485–494, doi:10.1016/j.chaos.2017.09.010.
[46] A. Zareie, A. Sheikhahmadi, M. Jalili, Identification of influential users in social networks based on users' interest, Inf. Sci. 493 (2019a) 217–231, doi:10.1016/j.ins.2019.04.033.

[47] A. Zareie, A. Sheikhahmadi, M. Jalili, Influential node ranking in social networks based on neighborhood diversity, Future Gener. Comput. Syst. Int. J. Esci. 94 (2019b) 120–129, doi:10.1016/j.future.2018.11.023.
[48] A. Zareie, A. Sheikhahmadi, K. Khamforoosh, Influence maximization in social networks based on TOPSIS, Expert Syst. Appl. 108 (2018) 96–107, doi:10.1016/j.eswa.2018.05.001.
[49] S.L. Zhang, M. Medo, L.Y. Lu, M.S. Mariani, The long-term impact of ranking algorithms in growing networks, Inf. Sci. 488 (2019a) 257–271, doi:10.1016/j.ins.2019.03.021.
[50] W. Zhang, J. Yang, X.Y. Ding, X.M. Zou, H.Y. Han, Q.C. Zhao, Groups make nodes powerful: identifying influential nodes in social networks based on social conformity theory and community features, Expert Syst. Appl. 125 (2019b) 249–258, doi:10.1016/j.eswa.2019.02.007.