

Identifying influential nodes based on fuzzy local dimension in complex networks

Tao Wen, Wen Jiang*

School of Electronics and Information, Northwestern Polytechnical University, Xi'an, shaanxi 710072, China

ARTICLE INFO

Article history:

Received 27 January 2018

Accepted 11 January 2019

Keywords:

Complex networks
Influential nodes
Local dimension
Fuzzy sets
Centrality measure

ABSTRACT

How to identify influential nodes in complex networks is an important aspect in the study of complex network. In this paper, a novel fuzzy local dimension (FLD) is proposed to rank the influential nodes in complex networks, where a node with high fuzzy local dimension has high influential ability. This proposed method focuses on the influence of the distance from the center node on the local dimension of center node by fuzzy set, resulting in a change in influential ability. In order to show this proposed method's effectiveness and accuracy, four real-world networks are applied in this paper. Meanwhile, Susceptible-Infected (SI) is used to simulate the spreading process by FLD and other centrality measures, and Kendall's tau coefficient is used to describe the correlation between the influential nodes obtained by centrality and the results measured by SI model. Observing from the ranking lists and simulated results, this method is effective and accurate to rank the influential nodes.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

In recent researches, a great deal of systems in the real world exist in the form of networks, such as the relationship network in the social system [1,2], protein network [3,4], scientists collaboration network [5]. Because the complex network [6] can show the structural complexity and the original properties of the original system, it is developing rapidly and widely used in many natural sciences over the years. The study of complex network can help to quantify the fuzzy world, predict the development of the system [7,8], and study the prisoner's dilemma experiments [9–11]. Complex networks have produced a large number of practical models and these have achieved a lot of practical results [12–14].

The relationship between two nodes like diseases spreading [15,16] and information dissemination [17] in real-world networks has important significance [18]. If a node has greater influence in complex network, this node undertakes the heavier task in the exchange of information (the wider the scope of information diffusion from this node within the unit time), the greater the impact of removing this node in the network. Based on this, identifying influential nodes in complex network has been focus of the research of complex networks, which can control the spread of disease [19] and identify the most influential spreaders [20]. A great deal of centrality measures of identifying the influential

nodes have been proposed in the past years [21,22]. The first person to identify the influential nodes is Shimbel [23] in 1953, and the method is that the node's centrality is the number of shortest paths which go through the selected node. Then, many methods with high accuracy are proposed to identify the influential nodes like Betweenness centrality (BC) [24,25], Degree centrality (DC) [24], Closeness centrality (CC) [24], Eigenvector centrality (EC) [26], Local dimension (LD) [27] and so on. Degree centrality (DC) is a simple method which doesn't pay attention to the global structure, so it's lack of accuracy. Betweenness centrality (BC) and Closeness centrality (CC) focus on the global structure and are more accurate comparing DC, but the computational complexity limits its application to large scale real-world complex networks. Meanwhile, CC can't be applied in the networks with disconnected nodes which means there is no shortest path between two nodes. Eigenvector centrality (EC) has greater limitations because it can't be applied to asymmetric networks, and many positions can't be chosen. Local dimension (LD) focus on the nodes whose shortest distance from the center node are less than the box-size equally. But in general, the node contribution is different from each other because of the different distance from the center node. So this method isn't accurate. Many measures [28] like semi-local centrality (SLC) [29], LeaderRank (LR) [30,31] and the role of neighborhood [32] have been proposed and the results show the efficient performance of these methods.

Fuzzy theory is proposed by L. A. Zadeh [33] in 1965, and fuzzy set can be more accurate description of information [34]. Because of it's effective to describe the uncertain information, fuzzy sets

* Corresponding author.

E-mail address: jiangwen@nwpu.edu.cn (W. Jiang).

has been applied in many subjects [35] like Ant Colony Optimization [36,37], volatility forecasting [38,39], decision making [40,41], supplier selection [42], similarity measure [43], information granules [44,45], designing type-2 fuzzy systems [46,47], uncertainty measure [48,49] and combination of D-S theory [50,51]. The results show the effectiveness and accuracy of this method. The fuzzy sets has already been applied to calculate the fractal dimension by Pedryca [52], and Castillo et al. apply the fuzzy fractal dimension in many fields, such as time series prediction [53], measuring the complexity of the sound signal [54]. Recently, some extensions of fuzzy set are presented to model more complicated situations like D numbers [55–57]. Due to the uncertainty in complex networks, fuzzy fractal dimension of complex network is proposed by Zhang et al. [58], and it has been applied in analysing images [59], measuring financial data [60] and dealing with big data [61]. It can describe the covering ability of box more accurate and have less time consuming. Fractal dimension can depict the self-similarity and fractal properties of complex networks [62], so it has been applied in many fields [63–65]. The local dimension has different values with different center nodes and different radius, and it's a great progress in dimension of complex networks. Because of the difference of the node's local dimension, Pu et al. [27] applied it to identify the influential nodes, and the ranking list of influential nodes is similar to the previous centrality measures. In this method, the nodes in the box are considered equally to the local dimension which is not precise.

In this paper, a novel centrality measure of identifying influential nodes is proposed which is based on fuzzy local dimension (FLD). The fuzzy sets can detailed describe the relationship between two nodes based on the topological distance between this two nodes, which can describe the topological structure more accurate. The dimension of complex networks can reveal the topological structure and dynamics structure which both have important influence to rank the influential nodes. Based on these, fuzzy local dimension is proposed to identify the influential nodes, and this method is applicable to most complex networks and takes less time than some measures like BC. The larger the fuzzy local dimension of a node, the greater the impact of this node, the wider the scope of information diffusion from this node within the unit time. In order to evaluate the validity and reliability of this method, four real-world complex networks have been applied in this paper to compare this method with other centrality measures. Susceptible-Infected (SI) [66] model is used to simulate the spreading progress by different centrality measures, and Kendalls tau coefficient is used to measure the correlation between the influential nodes ranking list obtained by centrality measures including fuzzy local dimension and the result measured by SI model. The results show this proposed method performs well.

The remainder of this paper is organized as follows. Some brief overview of complex network, fuzzy sets and centrality measures are given in Section 2. In Section 3, the method which is based on the fuzzy local dimension to identify the influential nodes is proposed. Susceptible-Infected (SI) and Kendall's tau coefficient is used to evaluate the results obtained by FLD in Section 4. Some conclusions are given in Section 5.

2. Preliminaries

In complex networks $G(N, V)$, there are a collection of nodes $N = (1, 2, \dots, n)$ and a collection of edges $V = (1, 2, \dots, v)$.

2.1. The shortest path between any two nodes

The shortest path between any two nodes can influence the properties of the complex network. The adjacency matrix of complex network G can be obtained by the relationship between nodes

and edges, which can be represented by $X = (x_{ij})$. The case when $x_{ij} = 1$ represents there is a connection between node i and node j , and $x_{ij} = 0$ is the opposite case. The shortest distance [67] between node i and node j is expressed as d_{ij} , and the maximum value of shortest distance from center node i is expressed as d_i^{\max} . It should be noted that the short path finding issue plays a very important role in network optimization [68,69].

2.2. Centrality measures

There are several node centrality measures that already exist, like degree centrality (DC), closeness centrality (CC), betweenness centrality (BC), eigenvector centrality (EC), local dimension (LD) which are defined as follows,

The DC [6] is the degree centrality of node i which is expressed as $C_D(i)$. It equals to the number of edges connected with this center node.

Because the degree centrality only considers the direct influence of node, without considering the influence of network structure to the information disseminating, the closeness centrality (CC) is proposed by Freeman which is expressed as $C_C(i)$. In general, information would disseminate in the shortest distance. If a node with the ability to forward the most of the information is lost, the network transmission efficiency would have a large impact. The more shortest distance going through node i , the more importance node i .

Definition 2.1. ((CC) [24]). The closeness centrality $C_C(i)$ is defined as follows,

$$C_C(i) = \frac{1}{\sum_{j=1}^N d_{ij}} \quad (1)$$

where node i is the center node, j is the other nodes in the whole network, N is the total number of nodes in this network, d_{ij} is shortest distance from node i to node j .

Because the closeness centrality (CC) has high computational complexity which can't deal with the large scale network well, betweenness centrality (BC) is proposed by Freeman which is expressed as $C_B(i)$. The smaller the sum of the shortest distance d_{ij} from center node i to all other nodes in the network, the less time for information disseminating, and the more influential the node i .

Definition 2.2. ((BC) [25]). The betweenness centrality $C_B(i)$ is defined as follows,

$$C_B(i) = \frac{\sum_{s \neq t \neq i} L_{st}(i)}{\sum_{s \neq t} L_{st}} \quad (2)$$

where L_{st} is the number of shortest paths between node s and node t in the whole network, $L_{st}(i)$ is the number of shortest paths between node s and node t which go through node i .

Definition 2.3. ((EC) [26]). The EC is the eigenvector centrality of node i which is expressed as $C_E(i)$. A is an $n \times n$ similarity matrix, and the eigenvector centrality x_i of node i is the i th entry in the normalized eigenvector which belongs to the similarity matrix A .

$$Ax = \lambda x, x_i = u \sum_{j=1}^n a_{ij} x_j, i = 1, 2, \dots, n \quad (3)$$

where λ is the largest eigenvalue of similarity matrix A , n is the number of nodes, and the proportionality which is defined as follows,

$$u = \frac{1}{\lambda}$$

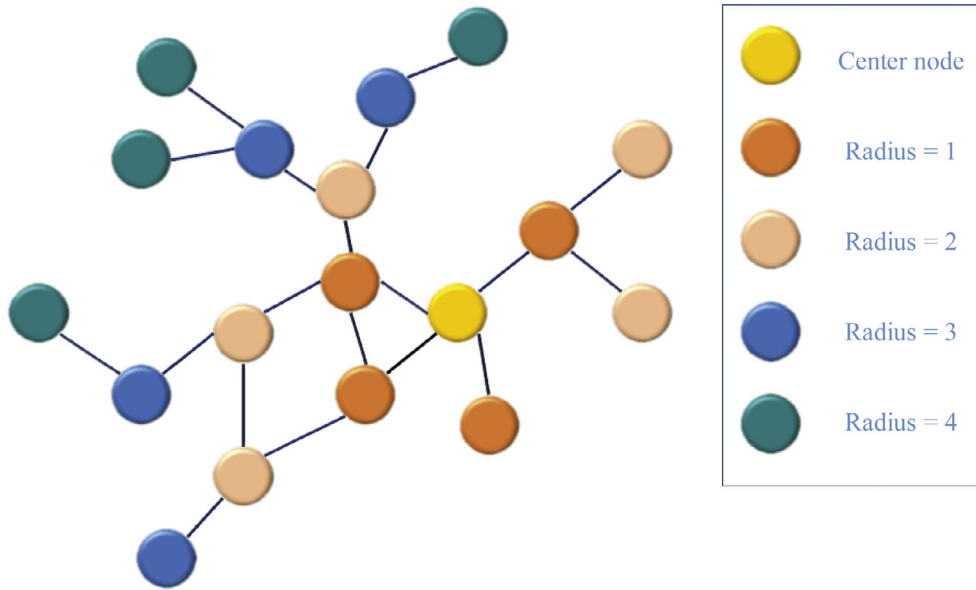


Fig. 1. The example network.

so x_i is the sum of similarity scores of the nodes which are connected to this node.

The local dimension $C_{LD}(i)$ [27] pays attention to the individuals rather than only on the global information, and it's introduced detailed in the next section.

2.3. Local dimension of complex networks

To express the local properties of each node in the complex network, Silva et al. have proposed local dimension. This method can observe the different scaling properties with different topological distance r from the selected center node, and it has been proved that not only the small-world networks but also many real complex networks follows a power law distribution. It means that the radius r (it can equal to r) and the total number of nodes $B_i(r)$ within the radius r for each node i have a power law relation as follows,

$$B_i(r) \sim \mu r^{D_i} \quad (4)$$

where D_i represents the local dimension of node i . It can be obtained by the slope of the double logarithmic scale fitting curves and Eq. (4) can be rewritten as follows,

$$D_i = \frac{d}{d \log r} \log B_i(r) \quad (5)$$

where the radius r is expanding from 1 to the maximum value of shortest distance d_i^{\max} from center node i , so the radius is discontinuous. Because of the discrete nature [70,71] in complex network, the derivative of Eq. (5) is still valid in this case and is expressed as follows,

$$D_i = \frac{r}{B_i(r)} \frac{dB_i(r)}{dr} \quad (6)$$

$$D_i \approx r \frac{n_i(r)}{B_i(r)} \quad (7)$$

where $n_i(r)$ is the number of nodes whose shortest distance d_{ij} from node i is equal to radius r , $B_i(r)$ is the number of nodes whose shortest distance d_{ij} from node i is less than or equal to radius r . The local dimension $C_{LD}(i)$ of node i is the slope of the double logarithmic scale fitting curves ($\log B_i(r_t)$ vs $\log r_t$) by liner regression.

The smaller the local dimension $C_{LD}(i)$, the more importance the node i .

Example 2.1. A simple network can explain how local dimension can be obtained. The network is shown as Fig. 1, and the max value of shortest distance d_i^{\max} from the selected center node can be obtained by its topological structure. The relationship between the number of nodes $B_i(r)$ within radius and the radius r is shown in Fig. 2. The local dimension of the center node is the slope of the liner regression in the double logarithmic scale.

2.4. Fuzzy sets

In the traditional case of things divided into two categories, when there is a class C which is a subset of the universal set X , any case of an input variable $x \in X$ whether belongs to the given subset C or not. There is a characteristic function $I_C(x) \rightarrow \{0, 1\}$ which is defined as follows,

$$I_C(x) = \begin{cases} 0, & x \in C \\ 1, & x \notin C \end{cases} \quad (8)$$

where there is a premise $x \in X$.

Facing real world situations, there is no clear boundary between two categories or the boundary may be overlapping. So it's uncertain that the input variable x belongs to the subset C totally. To deal with this problem, this characteristic function must be improved to describe the intermediate value between 0 and 1. L. A. Zadeh [33] proposed the fuzzy sets to modify the characteristic function $I_C(x)$ to the membership function $\mu_C(x)$ which can describe the interval continuous function between 0 and 1.

3. Fuzzy local dimension of complex networks

3.1. Basic method

Based on the fuzzy sets and local dimension, a new centrality measure called fuzzy local dimension is proposed. Local dimension focus on the nodes equally whose shortest distance from the center node are less than the box-size. But in general, the node's contribution is different from each other because of the different distance from the center node. It can be easily obtained that the smaller the distance, the greater the contribution to the center

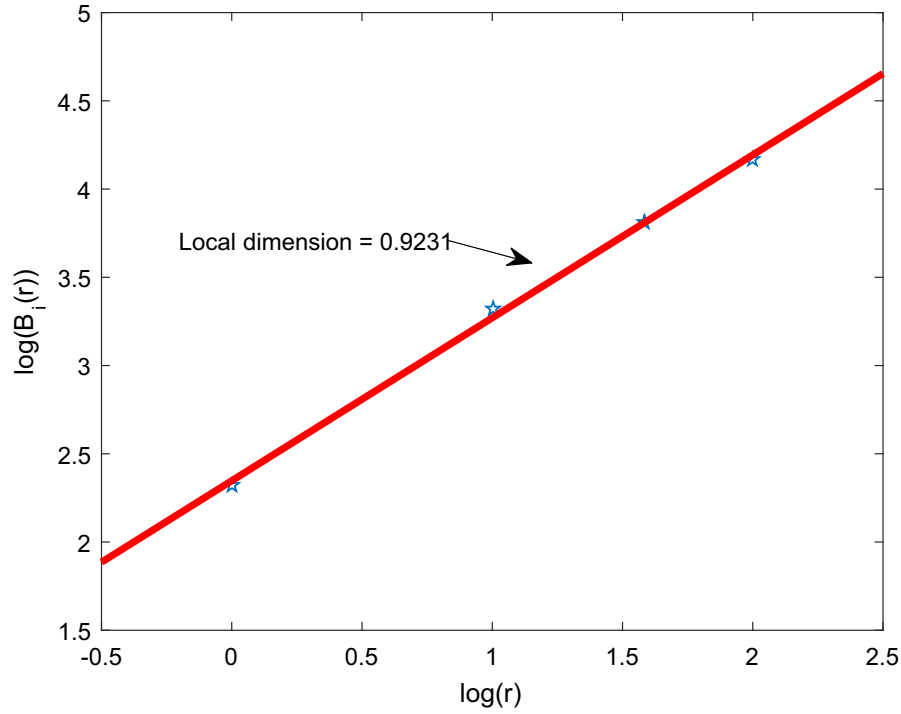


Fig. 2. The relationship in example network.

node. To improve this model, fuzzy sets are used to distribute between nodes contributing to the local dimension, and each node within the box-size has a weight through the fuzzy membership function. So the fuzzy local dimension (FLD) is proposed in this paper.

Definition 3.1. (Fuzzy local dimension). Like the local dimension D_i , the fuzzy local dimension D_{fuzzy_i} is also focus on the local properties of each node, which has the same expression as the local dimension D_i and is defined as follows,

$$D_{fuzzy_i} = \frac{d}{d \log r_t} \log N_i(r_t, \varepsilon) \quad (9)$$

because of the discrete nature [70,71] in complex network, the derivative of Eq. (9) is still valid in this case and is expressed as follows,

$$D_{fuzzy_i} = \frac{r_t}{N_i(r_t, \varepsilon)} \frac{d}{dr_t} N_i(r_t, \varepsilon) \quad (10)$$

$$D_{fuzzy_i} \approx r_t \frac{n_i(r_t)}{N_i(r_t, \varepsilon)} \quad (11)$$

where r_t is the radius from center node i and it can expand from small to big in a set $\{1, 2, \dots, d_i^{\max}\}$, $n_i(r_t)$ is the fuzzy number of nodes whose shortest distance equal to the box-size ε , $N_i(r_t, \varepsilon)$ is the fuzzy number of nodes whose shortest distance is less than the box-size ε which is obtained by a fuzzy set, and it can be defined as follows,

$$N_i(r_t, \varepsilon) = \frac{\sum_{j=1}^N A_{ij}(\varepsilon)}{N_{i_r}} \quad (12)$$

where ε is the size of box, N_{i_r} is the real number of nodes when the shortest distance between node i and j is less than the box size ε , and $A_{ij}(\varepsilon)$ is a membership function when the distance from node j to node i is less than the box size ε which is defined as

follows,

$$A_{ij}(\varepsilon) = \exp\left(-\frac{d_{ij}^2}{\varepsilon^2}\right) \quad (13)$$

where d_{ij} is the shortest distance between center node i and node j . $A_{ij}(\varepsilon)$ follows a normal distribution, it can give the neighbour node which is relative to the center node i a weight between 0 and 1 instead of a definite value of 0 or 1. It's a function of distance and it can distinguish the different nodes whose distance is different from the center node. When the nodes have different distance with center node, these nodes would have different contributions to the center node. When the neighbour node is close to the center node, it would have a bigger weight which is close to 1, and a node's weight would be a smaller value which is close to 0 when the neighbour node away from the center node. The large the membership function $A_{ij}(\varepsilon)$ for node i , the greater the contribution to the fuzzy number of nodes $N_i(r_t, \varepsilon)$, and node i would be more influential.

With the membership function $A_{ij}(\varepsilon)$, the node's contribution is not equal to the center node any more. Each node has a individual weight which is relative to the shortest distance from the center node. The closer the node is to the center node, the greater the impact on the center node, the greater the fuzzy number of nodes $N_i(r_t, \varepsilon)$, the more influential the center node. This proposed method is more realistic and effective. When a node is very unimportant, this node's fuzzy local dimension may be negative, and this property gives a critical basis for identifying the influential nodes.

3.2. Example explanation

In order to explain this proposed method more specific, a small real world complex network named as Kite network [72] is used to explain how this method works. This small network is a interpersonal relation network with ten nodes which is shown in Fig. 3. The Kite network describes the relationships between 10 different

Table 1

The fuzzy local dimension of each node in Kite network.

Node ID	1	2	3	4	5
Fuzzy local dimension	0.3609	0.3609	0.3015	0.4554	0.4554
Node ID	6	7	8	9	10
Fuzzy local dimension	0.3015	0.4442	0.0760	0.0375	-0.1163

people, and it consists of two parts. People in left part (node 1, 2, 3, 4, 5, 6, and 7) are close to each other, and people in right part (node 8, 9, and 10) are not close.

Step 1: All the nodes in this network will be selected as the center node to calculate its fuzzy local dimension in turn. In this example, node 7 is chosen as the center node first.

Step 2 : The fuzzy number of nodes $N_i(r_t)$ with corresponding radius r_t can be obtained by Eq. (12), where the radius r_t would expand from 1 to the maximum value of shortest distance d_7^{\max} from the center node 7. The fuzzy number of nodes $N_7(r_t)$ when the radius r_t equals to 1, 2, 3, and 4 can be respectively obtained as follows,

$$\begin{aligned}
 N_7(1) &= \frac{\exp\left(-\frac{d_{17}^2}{2^2}\right) + \exp\left(-\frac{d_{27}^2}{2^2}\right) + \exp\left(-\frac{d_{37}^2}{2^2}\right) + \exp\left(-\frac{d_{47}^2}{2^2}\right) + \exp\left(-\frac{d_{57}^2}{2^2}\right) + \exp\left(-\frac{d_{67}^2}{2^2}\right) + \exp\left(-\frac{d_{77}^2}{2^2}\right)}{7} \\
 &= \frac{\exp(-1) + \exp(-1) + \exp(-1) + \exp(-1) + \exp(-1) + \exp(-1) + \exp(0)}{7} \\
 &= \frac{6 \times \exp(-1) + \exp(0)}{7} \\
 &= 0.4582
 \end{aligned}$$

$$\begin{aligned}
 N_7(2) &= \frac{\exp\left(-\frac{d_{17}^2}{2^2}\right) + \exp\left(-\frac{d_{27}^2}{2^2}\right) + \exp\left(-\frac{d_{37}^2}{2^2}\right) + \exp\left(-\frac{d_{47}^2}{2^2}\right) + \exp\left(-\frac{d_{57}^2}{2^2}\right) + \exp\left(-\frac{d_{67}^2}{2^2}\right) + \exp\left(-\frac{d_{77}^2}{2^2}\right) + \exp\left(-\frac{d_{78}^2}{2^2}\right)}{8} \\
 &= \frac{\exp\left(-\frac{1}{4}\right) + \exp\left(-\frac{1}{4}\right) + \exp\left(-\frac{1}{4}\right) + \exp\left(-\frac{1}{4}\right) + \exp\left(-\frac{1}{4}\right) + \exp\left(-\frac{1}{4}\right) + \exp(0) + \exp(-1)}{8} \\
 &= \frac{6 \times \exp\left(-\frac{1}{4}\right) + \exp(0) + \exp(-1)}{8} \\
 &= 0.7551
 \end{aligned}$$

$$\begin{aligned}
 N_7(3) &= \frac{6 \times \exp\left(-\frac{1}{3^2}\right) + 1 \times \exp\left(-\frac{2^2}{3^2}\right) + 1 \times \exp\left(-\frac{3^2}{3^2}\right) + 1 \times \exp(0)}{9} \\
 &= 0.8198
 \end{aligned}$$

$$\begin{aligned}
 N_7(4) &= \frac{6 \times \exp\left(-\frac{1}{4^2}\right) + 1 \times \exp\left(-\frac{2^2}{4^2}\right) + 1 \times \exp\left(-\frac{3^2}{4^2}\right) + 1 \times \exp\left(-\frac{4^2}{4^2}\right) + 1 \times \exp(0)}{10} \\
 &= 0.8353
 \end{aligned}$$

Step 3 : The relationship between fuzzy number of nodes $N_7(r_t)$ and the corresponding radius r_t can be shown in Fig. 4. The slope of the double logarithmic scale fitting curves is the fuzzy local dimension $D_{fuzzy,7}$ of node 7.

Step 4 : Repeat step 1 to step 3 for all remaining nodes, and the results are shown in Table 1.

Finally, the nodes would be sorted by the value of fuzzy local dimension, and the ranking results with this proposed method and other centrality measures are shown in Table 2. Observing from Tables 1 and 2, node 4(Fernando) and node 5(Garth) play the most important roles at the same time because they are equivalent to the door of the parts of the network, and these two nodes' fuzzy local dimension are also maximum. These two nodes connect two

Table 2

Kite network ranking list with different centrality measures.

Rank	BC	CC	DC	EC	LD	FLD
1	8	4	7	7	7	4
2	4	5	5	5	4	5
3	9	7	4	4	5	7
4	5	8	2	1	1	2
5	7	2	1	2	2	1
6	2	1	8	6	3	6
7	1	6	6	3	6	3
8	10	3	3	8	9	8
9	6	9	9	9	10	9
10	3	10	10	10	8	10

parts of the network. The left part of the network is more complex, so the nodes in this part are more important. Meanwhile, in the left part of Kite network, node 7(Diane) is the center of this part, so it is the most influential node in the left half of the network. The rest nodes in the left part of network including node 1(Beverly), node 2(Andre), node 3(Carol), node 6(Ed) play the same role

and they are the edges of this part, the presence of these nodes makes the left part of the network more connected, so they have the same importance. Then, the right part is less important in this network, so these three nodes rank in the last three in the ranking list. But node 8(Heather) is nearest to node 4 and node 5 which are the connection nodes, node 9(Ike) is closer than node 10(Jane), so their importance ranking is 8, 9, 10. This order is the same as the ranking lists obtained by FLD, so the final ranking results from this proposed method (FLD) is more reasonable than others. A detail information in Table 1 is that the fuzzy local dimension of node 10 is negative, and the above conclusion shows node 10 is the least

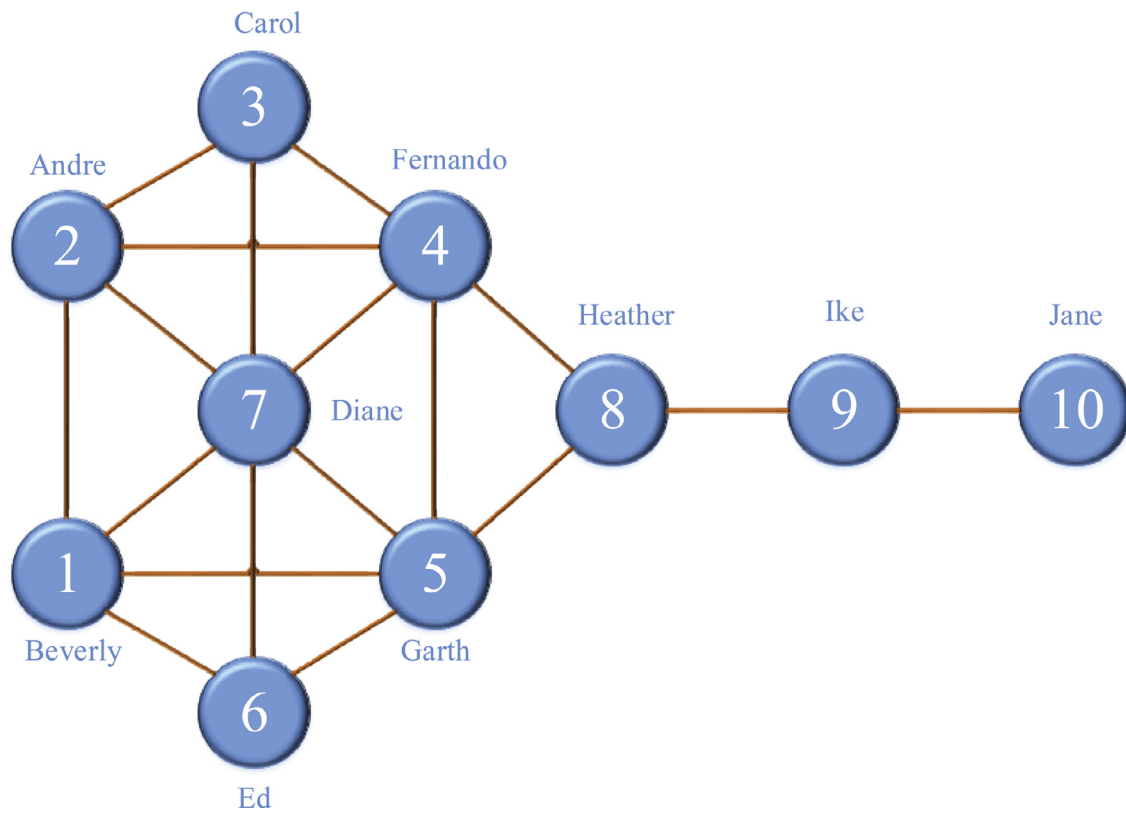


Fig. 3. The Kite network.

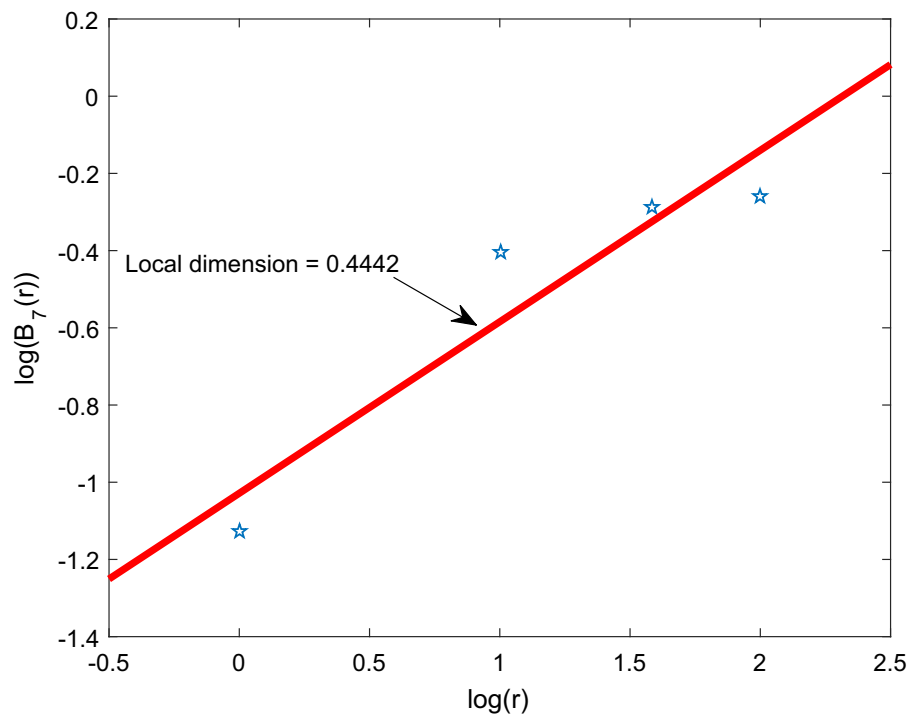


Fig. 4. The double logarithmic scale fitting curves of node 7 in Kite network.

Table 3

Topological properties in these four real-world complex networks. Nodes and edges represent the number of nodes and edges in these complex networks respectively; $\langle k \rangle$ and k_{\max} show the average and maximum value of degree; $\langle d \rangle$ and d_{\max} express the average and maximum value of shortest distance; C is the clustering coefficient of these networks.

Network	nodes	edges	$\langle k \rangle$	k_{\max}	$\langle d \rangle$	d_{\max}	C
email	1133	10,902	9.6222	71	3.6028	8	0.2202
USAir	332	2126	12.8072	139	2.9299	6	0.6252
karate	34	78	4.5882	17	2.3374	5	0.6175
Groad	1168	2486	2.1284	12	19.4024	62	0.0012

Table 4

Email network.

Rank	BC	CC	DC	EC	LD	FLD
1	23	333	105	105	105	333
2	105	23	333	16	333	23
3	333	105	42	196	23	42
4	76	42	23	204	42	105
5	42	41	16	42	16	76
6	578	76	41	49	434	468
7	135	233	196	56	41	41
8	41	52	233	116	14	233
9	52	135	76	333	468	52
10	355	378	21	3	299	378

Table 5

USAir network.

Rank	BC	CC	DC	EC	LD	FLD
1	118	118	118	118	118	118
2	8	261	261	261	261	261
3	261	67	255	255	152	67
4	47	255	182	182	230	255
5	201	201	152	152	255	201
6	67	182	230	230	182	182
7	313	47	166	112	112	166
8	13	248	67	67	147	47
9	182	166	112	166	166	248
10	152	112	201	147	293	112

important node of Kite network. So this information shows that the fuzzy local dimension would be negative when one node is far less important than other nodes in the network.

4. Experimental study

In order to test the effectiveness and accuracy of this proposed method, four real-world complex networks have been applied for this proposed method and some existing methods such as DC, CC, BC, EC, LD. These real-world complex networks include USAir network which can be download from (<http://vlado.fmf.uni-lj.si/pub/networks/data/>), email network which is from (<http://www.cs.bris.ac.uk/steve/peacockpaper/>), karate network which is from (<http://vlado.fmf.uni-lj.si/pub/networks/data/>), and Groad network which is from Ref. [73]. These networks' topological properties are shown in Table 3 respectively.

4.1. Data

BC, DC, CC, EC, LD and this proposed method (FLD) are used for four real-world complex networks to show the performance of this proposed method. The top-10 ranking lists of four real-world complex networks is shown in Tables 4–7. Observing from Table 4 for email network, the proposed method has the same nine node IDs with CC; there are seven same node IDs for the proposed method (FLD) and BC, DC. In Table 5 for USAir network, there are ten same node IDs for the proposed method and CC; FLD has the same

Table 6

karate network.

Rank	BC	CC	DC	EC	LD	FLD
1	1	1	34	34	34	1
2	3	3	1	1	1	34
3	34	34	33	3	33	33
4	33	32	3	33	24	3
5	32	33	2	2	3	2
6	6	14	32	9	2	32
7	2	9	4	14	30	24
8	28	20	24	4	6	28
9	24	2	14	32	7	31
10	9	4	9	31	28	30

Table 7

Groad network.

Rank	BC	CC	DC	EC	LD	FLD
1	219	698	693	219	219	698
2	543	219	403	217	369	450
3	758	450	300	267	217	565
4	693	565	758	198	565	219
5	886	331	410	207	693	763
6	698	763	373	331	267	267
7	735	267	217	565	295	663
8	403	729	556	236	207	331
9	565	663	331	295	758	295
10	763	347	219	231	331	729

nine node IDs with DC; there are six same node IDs for this proposed method and BC. In Table 6 for karate network, the proposed method (FLD) has the same eight node IDs with BC; FLD has the same seven and six node IDs with DC and CC. In Table 7 for Groad network, there are nine same node IDs for FLD and CC; there are some differences between this proposed method and BC, DC, and this proposed method has four and two same node IDs with BC and DC respectively. In Tables 6 and 7, there is a difference between FLD and EC. That's because these two methods have their own emphasis on the ranking results. EC pays more attention to the whole structure of complex networks, but FLD focuses on the each individuality rather than the whole structure of complex networks. Even if their respective priorities are different, the first few nodes of the ranking result are the same, and the difference of the ranking results obtained by FLD and other centrality measures is relative small. These show that FLD can get a relatively correct influential ranking lists.

This proposed method is better than LD because there are more same nodes with other centrality measures like BC, DC, CC. In email network, FLD has two and four more same node IDs than LD with BC, CC respectively; this proposed method has one more node ID than LD with DC. In USAir network, this proposed method has four and two more same node IDs than LD with CC, BC respectively; this proposed method has equal same node ID with LD comparing with DC. In karate network, FLD has one more node ID than LD with BC, CC, DC. In Groad network, FLD has four more same node IDs than LD with CC, this proposed method has equal same node ID with LD comparing with BC, and there are one less node ID than LD with DC. From the above results, a conclusion can be easily obtained that this proposed method (FLD) is more similar than LD comparing the existed centrality measures.

4.2. Measuring effectiveness by SI model

The standard of susceptible-infected (SI) model is used to test the different centrality measures result and capture the nodes' influential rank list. In SI model, these nodes are separate into two parts: (1) Susceptible, the number of susceptible nodes at time t is represented by $S(t)$. If one node is susceptible to this disease,

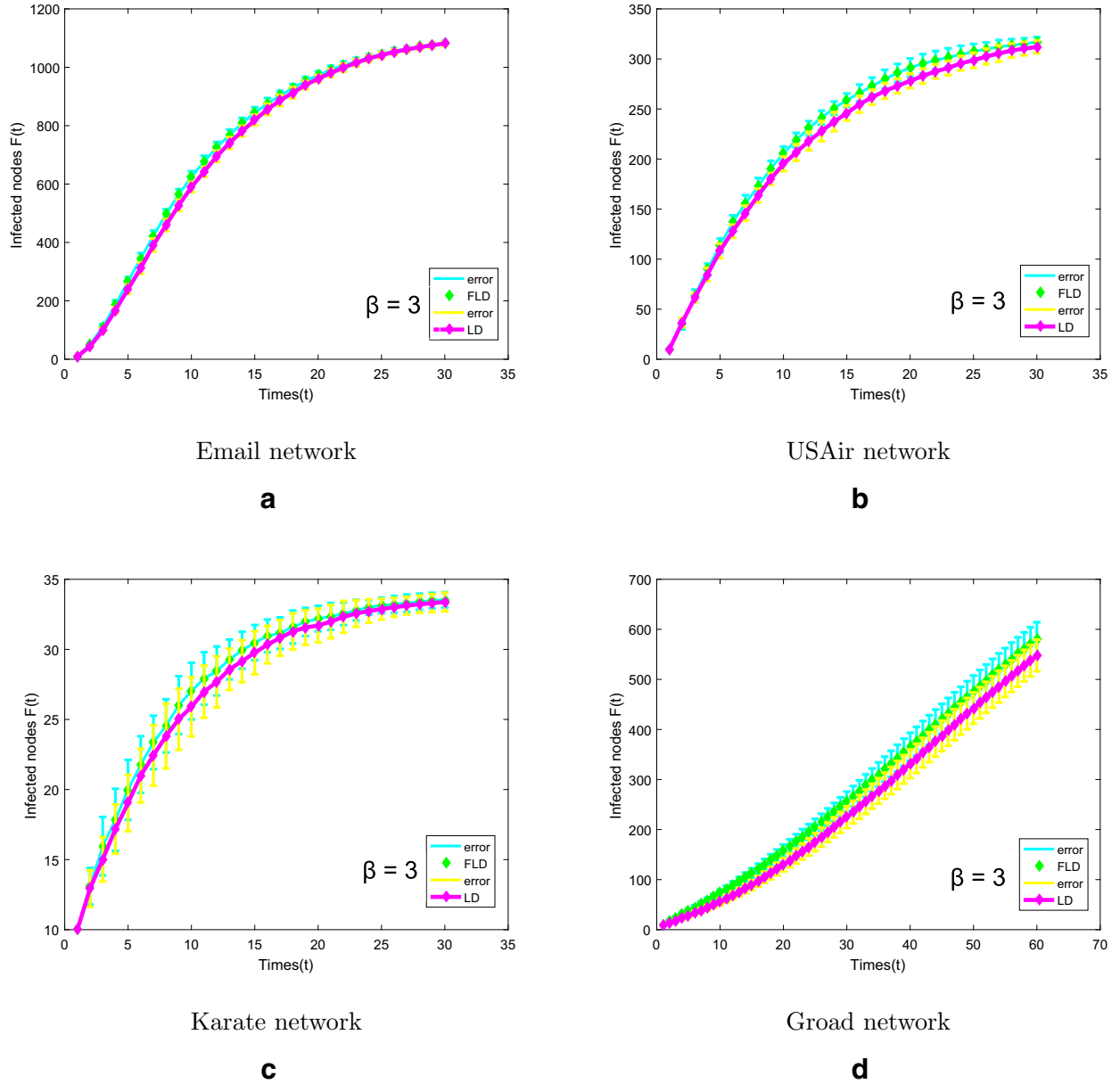


Fig. 5. The number of infected nodes with different initial infected nodes (Top-10 nodes in FLD and LD) for four real-world complex networks when $\beta=3$.

this node would be affected by the neighbor nodes to become infected in the next moment. (2) Infected, the number of infected nodes at time t is represented by $I(t)$. If a node is infected, this node has the ability to affect the neighbor nodes. The initial infected nodes $I(0)$ are the test nodes which have been already gotten by previous centrality measures or random given nodes. It can be easily obtained that $S(t) + I(t) = N$, where N the total number of nodes in complex networks. At each time t , every infected nodes have the spreading rate λ which is defined as $\lambda = (\frac{1}{2})^\beta$ to infect their neighbor susceptible nodes, where β has different values in different complex networks. After infection, the node in S state would change to I state. In general, the more influential a node, the stronger its ability to spread, the more the number of nodes infected in a certain time, the more influential this node. The number of infected nodes also can be denoted as $F(t)$, which can be an indicator to measure the importance of the initial given infected nodes at time t . It can be easily obtained that $F(t)$ would increase with the development of time t , and $F(t)$ would be constant when

all nodes have been infected, here it's denoted as $F(t_e)$, where t_e is the final time when there isn't node to be infected. Obviously, the higher $F(t)$, the more influential the initial nodes, and $F(t_e)$ represents the maximum influence of initial nodes.

According to the property of SI model, a node which has stronger spreading ability is more important in the network. In order to compare the spreading ability of the top-10 nodes between this proposed method and local dimension which are shown in Tables 4–7, SI model is used to simulate the spreading progress of the top-10 nodes and the results of average of 100 independent tests are shown in Fig. 5. After the same time t , if the number of infected node $F(t)$ obtained by FLD is greater than LD, it can be shown that the spreading ability of the top-10 nodes obtained by FLD is stronger than obtained by LD, and the opposite case is also true. In Fig. 5, the larger the number of infected node $F(t)$, the larger the vertical value is, the higher the curve is. In order to ensure the accuracy of the experiment and to prevent contingency, the simulation process with different top-10 infected nodes would

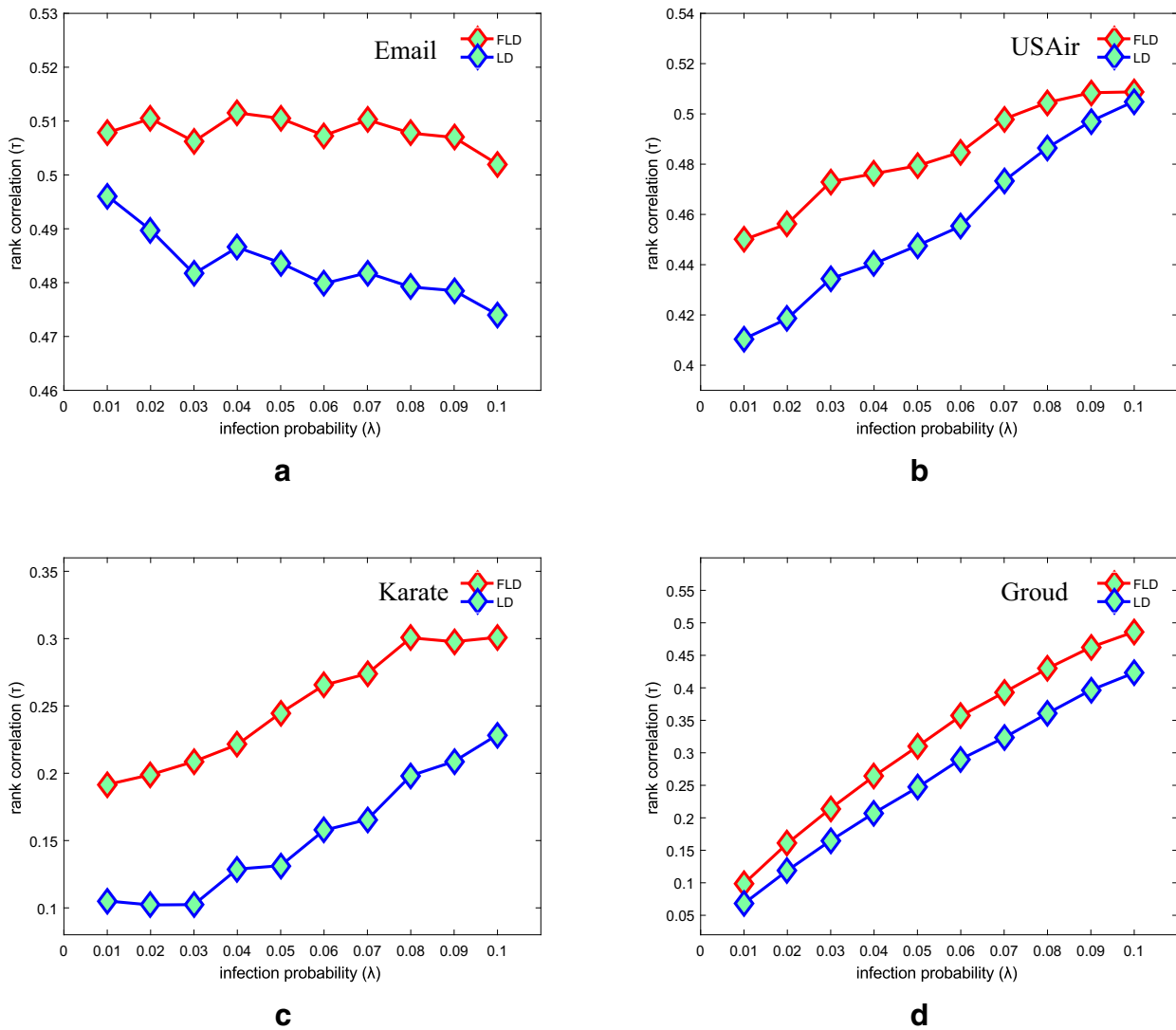


Fig. 6. The Kendall's tau coefficient τ is obtained by FLD, LD and the ranking list from SI model on different networks. The result is the average of 100 independent experiments and the spreading probability λ is increasing from 0.01 to 0.1.

be performed 100 times. So the curve represents the mean value of 100 experiments, and the error represents the possible interval at each time t .

For Email network shown in Fig. 5(a), the initial nodes in this proposed method have stronger spreading ability than LD, because the number of infected nodes $F(t)$ by FLD is larger than LD in the middle infecting progress. For USAir network shown in Fig. 5(b), the number of infected nodes $F(t)$ by this proposed method is larger than LD in whole progress, so this proposed method performs better than LD. For Karate network and Groud network shown in Fig. 5(c) and (d), although the error of average of 100 independent tests is larger than Email network and USAir network, the number of infected nodes $F(t)$ by FLD is larger than LD obviously. So this proposed method outperforms LD. Through these results, an obvious conclusion can be obtained that the initial nodes obtained by this proposed method (FLD) have stronger spreading ability than LD, so FLD is more useful to identify the influential nodes in complex networks.

4.3. Measuring effectiveness by Kendall's tau coefficient

To measure the correlation between the influential nodes obtained by centrality measures including fuzzy local dimension

and the result measured by SI model, Kendall's tau coefficient [74] which is applied in many fields [75] is used in this paper. The higher the Kendall's tau coefficient, the more similar the two sequences, the more accurate this proposed method.

Two random variables W and V are assumed, and their i th combination is represented as (W_i, V_i) . If both $W_i > W_j$ and $V_i > V_j$ or if both $W_i < W_j$ and $V_i < V_j$ occur at the same time, the i th combination (W_i, V_i) and j th combination (W_j, V_j) are considered concordant. If both $W_i > W_j$ and $V_i < V_j$ or if both $W_i < W_j$ and $V_i > V_j$ occur at the same time, it means that these two combinations are discordant. And if $W_i = W_j$ or $V_i = V_j$, these two combinations are neither concordant nor discordant.

Definition 4.1. (Kendall's tau coefficient) Then the Kendall's tau coefficient τ is defined [76,77] as follows,

$$\tau = \frac{n_c - n_d}{0.5n(n-1)} \quad (14)$$

where n is total combinations in these sequences, n_c and n_d are the number of concordant combinations and discordant combinations respectively. The higher the Kendall's tau coefficient τ , the more accurate the centrality measure which is used, and the smaller the τ is the opposite case. The case $\tau = 1$ represents the ranking list by

the centrality measures is same as the result obtained by SI model and the case $\tau = 0$ is the opposite case.

Kendall's tau coefficient τ can clearly show the similarity extent between two centrality measures, and it can use numerical results to intuitively show the correlation. When setting the spreading probability λ increasing from 0.01 to 0.1, the Kendall's tau coefficient τ would increase with the increasing of spreading probability and it is shown in Fig. 6. It's clear shown when the correlation between the centrality measures and SI model is higher, the Kendall's tau coefficient τ is larger. From Email network in Fig. 6(a), the Kendall's tau coefficient τ does not change greatly with the increasing of spreading probability, but the Kendall's tau coefficient τ by FLD is large than LD, so this proposed method outperforms LD. In USAir network and Karate network in Fig. 6(b) and (c), the Kendall's tau coefficient τ by FLD is much larger than LD, so this proposed method performs better than LD. For Groad network in Fig. 6(d), the Kendall's tau coefficient τ by FLD is larger than LD and the result shows that FLD outperforms LD. The comparison between Kendall's tau coefficient τ obtained by FLD and LD shows that this FLD outperforms LD and this proposed method is more useful to identify the influential nodes than LD.

5. Conclusion

In this paper, a novel fuzzy local dimension is proposed to rank the influential nodes in complex networks. In this method, the nodes with different values of shortest distance from the center node have a different weight to the local dimension of center node by membership function, which is more realistic. The larger the fuzzy local dimension of this center node, the higher ranking the node in the list, the more influential this node. The positive and negative change of the fuzzy local dimension gives a clear demarcation of the ranking list of influential nodes. Compared with other centrality measures, this proposed method focus on each individuality rather than the whole structure of complex networks. Compared with LD, this proposed method takes the different distances from the center node into consideration, and each node within the box-size has a weight through the fuzzy membership function. The node which is closer to the center node would have a greater impact on the center node. Applied to four real-world complex networks, the results from the proposed method (FLD) are in good agreement with the previous centrality measures, and it performs better than LD. Susceptible-Infected (SI) model is used to simulate the spreading progress by different centrality measures, and Kendall's tau coefficient is used to measure the correlation between the influential nodes obtained by centrality measures including fuzzy local dimension and the result measured by SI model. The results of these four real-world networks applied to these two models show the effectiveness and accuracy of this method.

Acknowledgment

The authors greatly appreciate the reviews' suggestions and the editor's encouragement. The work is partially supported by National Natural Science Foundation of China (Program No. 61671384, 61703338), Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2018JQ6085).

References

- [1] Deville P, Song CM, Eagle N, Blondel VD, Barabasi AL, Wang DS. Scaling identity connects human mobility and social interactions. *Proc Natl Acad Sci USA* 2016;113(26):7047–52. doi:10.1073/pnas.1525443113.
- [2] Wang DS, Song CM. Impact of human mobility on social networks. *J Commun Networks* 2015;17(2):100–9. doi:10.1109/jcn.2015.000023.
- [3] Zhao ZQ, Yu ZG, Anh V, Wu JY, Han GS. Protein folding kinetic order prediction from amino acid sequence based on horizontal visibility network. *Curr Bioinform* 2016;11(2):173–85. doi:10.2174/1574893611666160125221326.
- [4] Hahn K, Massopust PR, Prigarin S. A new method to measure complexity in binary or weighted networks and applications to functional connectivity in the human brain. *BMC Bioinform* 2016;17:18. doi:10.1186/s12859-016-0933-9.
- [5] Clough JR, Evans TS. What is the dimension of citation space? *Phys a-Stat MechAppl* 2016;448:235–47. doi:10.1016/j.physa.2015.12.053.
- [6] Newman MEJ. The structure and function of complex networks. *SIAM Rev* 2003;45(2):167–256. doi:10.1137/s003614450342480.
- [7] Gao JX, Barzel B, Barabasi AL. Universal resilience patterns in complex networks. *Nature* 2016;530(7590):307–12. doi:10.1038/nature16948.
- [8] Hong C, Zhang J, Cao X-B, Du W-B. Structural properties of the chinese air transportation multilayer network. *Chaos Solitons Fractals* 2016;86:28–34.
- [9] Li X, Jusup M, Wang Z, Li H, Shi L, Podobnik B, et al. Punishment diminishes the benefits of network reciprocity in social dilemma experiments. *Proc Natl Acad Sci* Published online 2017. doi:10.1073/pnas.1707505115.
- [10] Wang Z, Jusup M, Wang RW, Shi L, Iwasa Y, Moreno Y, et al. Onymity promotes cooperation in social dilemma experiments. *Sci Adv* 2017;3(3):e1601444.
- [11] Yu Y, Xiao G, Zhou J, Wang Y, Wang Z, Kurths J, et al. System crash as dynamics of complex networks. *Proc Natl Acad Sci* 2016;113(42):11726–31.
- [12] Zhang X, Feng L, Berman Y, Hu N, Stanley HE. Exacerbated vulnerability of coupled socio-economic risk in complex networks. *Epl* 2016;116(1):6. doi:10.1209/0295-5075/116/18001.
- [13] Duan SY, Wen T, Jiang W. A new information dimension of complex network based on r'enyi entropy. *Phys a-Stat Mech Appl* 2019;516:529–42. doi:10.1016/j.physa.2018.10.045.
- [14] Rosenberg E. Minimal partition coverings and generalized dimensions of a complex network. *Phys Lett A* 2017. doi:10.1016/j.physleta.2017.03.004.
- [15] Pei S, Makse HA. Spreading dynamics in complex networks. *J Stat Mech-TheoryExp* 2013. doi:10.1088/1742-5468/2013/12/p12002.
- [16] Wang Z, Bauch CT, Bhattacharyya S, d'Onofrio A, Manfredi P, Perc M, et al. Statistical physics of vaccination. *Phys Rep* 2016;664:1–113.
- [17] Jankowski J, Brodka P, Kazienko P, Szymanski BK, Michalski R, Rajdanowicz T. Balancing speed and coverage by sequential seeding in complex networks. *Sci Rep* 2017;7. doi:10.1038/s41598-017-00937-8.
- [18] Wen T, Song MX, Jiang W. Evaluating topological vulnerability based on fuzzy fractal dimension. *Int J Fuzzy Syst* 2018;20(6):1956–67. doi:10.1007/s40815-018-0457-8.
- [19] Fei L, Zhang Q, Deng Y. Identifying influential nodes in complex networks based on the inverse-square law. *Physica A* 2018;512:1044–59. doi:10.1016/j.physa.2018.08.135.
- [20] Morone F, Roth K, Min B, Stanley HE, Makse HA. Model of brain activation predicts the neural collective influence map of the brain. *Proc Natl Acad Sci USA* 2017;114(15):3849–54. doi:10.1073/pnas.1620808114.
- [21] Li M, Zhang Q, Deng Y. Evidential identification of influential nodes in network of networks. *Chaos Solitons Fractals* 2018;117:283–96.
- [22] Kang B, Deng Y, Hewage K, Sadiq R. A method of measuring uncertainty for Z-number. *IEEE Trans Fuzzy Syst* 2018. doi:10.1109/TFUZZ.2018.2868496.
- [23] Shimmel A. Structural parameters of communication networks. *Bull Math Biophys* 1953;15(4):501–7. doi:10.1007/BF02476438.
- [24] Freeman LC. Centrality in social networks conceptual clarification. *Soc Networks* 1979;1(3):215–39.
- [25] Newman MEJ. A measure of betweenness centrality based on random walks. *Soc Networks* 2005;27(1):39–54.
- [26] Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. *Soc Networks* 2001;23(3):191–201.
- [27] Pu J, Chen X, Wei D, Liu Q, Deng Y. Identifying influential nodes based on local dimension. *Epl* 2014;107(1). doi:10.1209/0295-5075/107/10010.
- [28] Bian T, Deng Y. Identifying influential nodes in complex networks: a node information dimension approach. *Chaos* 2018;28. doi:10.1063/1.5030894.
- [29] Chen D, Lu L, Shang M-S, Zhang Y-C, Zhou T. Identifying influential nodes in complex networks. *Phys a-Stat MechAppl* 2012;391(4):1777–87. doi:10.1016/j.physa.2011.09.017.
- [30] Li Q, Zhou T, Lu L, Chen D. Identifying influential spreaders by weighted leaderrank. *Phys a-Stat MechAppl* 2014;404:47–55. doi:10.1016/j.physa.2014.02.041.
- [31] Xue YK, Bogdan P. Reliable multi-fractal characterization of weighted complex networks: algorithms and implications. *Sci Rep* 2017;7:22. doi:10.1038/s41598-017-07209-5.
- [32] Liu Y, Tang M, Zhou T, Do Y. Identify influential spreaders in complex networks, the role of neighborhood. *Phys a-Stat MechAppl* 2016;452:289–98. doi:10.1016/j.physa.2016.02.028.
- [33] Zadeh L. Fuzzy sets. *Inf Control* 1965;8(3):338–53. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- [34] Pedrycz W, Wang X. Designing fuzzy sets with the use of the parametric principle of justifiable granularity. *IEEE Trans Fuzzy Syst* 2016;24(2):489–96. doi:10.1109/TFUZZ.2015.2453393.
- [35] Pedrycz W. From fuzzy data analysis and fuzzy regression to granular fuzzy data analysis. *Fuzzy Sets Syst* 2015;274:12–17. doi:10.1016/j.fss.2014.04.017.
- [36] Castillo O, Lizarraga E, Soria J, Melin P, Valdez F. New approach using ant colony optimization with ant set partition for fuzzy control design applied to the ball and beam system. *Inf Sci* 2015;294:203–15. doi:10.1016/j.ins.2014.09.040.
- [37] Castillo O, Neyoy H, Soria J, Melin P, Valdez F. A new approach for dynamic fuzzy logic parameter tuning in ant colony optimization and its application in

- fuzzy control of a mobile robot. *Appl Soft Comput* 2015;28:150–9. doi:10.1016/j.asoc.2014.12.002.
- [38] Pan L, Deng Y. A new belief entropy to measure uncertainty of basic probability assignments based on belief function and plausibility function. *Entropy* 2018;20(11):842. doi:10.3390/e20110842.
- [39] Maciel L, Ballini R, Gomide F. Evolving possibilistic fuzzy modeling for realized volatility forecasting with jumps. *IEEE Trans Fuzzy Syst* 2017;25(2):302–14. doi:10.1109/tfuzz.2016.2578338.
- [40] He Z, Jiang W. An evidential dynamical model to predict the interference effect of categorization on decision making. *Knowl Based Syst* 2018;150:139–49.
- [41] He Z, Jiang W. An evidential markov decision making model. *Inf Sci* 2018;467:357–72.
- [42] Deng X, Jiang W, Wang Z. Zero-sum polymatrix games with link uncertainty: a dempster-shafer theory solution. *Appl Math Comput* 2019;340:101–12.
- [43] Huang Z, Yang L, Jiang W. Uncertainty measurement with belief entropy on the interference effect in the quantum-like bayesian networks. *Appl Math Comput* 2019;347:417–28.
- [44] Pedrycz W, Jastrzebska A, Homenda W. Design of fuzzy cognitive maps for modeling time series. *IEEE Trans Fuzzy Syst* 2016;24(1):120–30. doi:10.1109/tfuzz.2015.2428717.
- [45] Pedrycz W. From fuzzy models to granular fuzzy models. *Int J Comput Intell Syst* 2016;9:35–42. doi:10.1080/18756891.2016.1180818.
- [46] Castillo O, Melin P. Optimization of type-2 fuzzy systems based on bio-inspired methods: a concise review. *Inf Sci* 2012;205:1–19. doi:10.1016/j.ins.2012.04.003.
- [47] Castillo O, Melin P, Pedrycz W. Design of interval type-2 fuzzy models through optimal granularity allocation. *Appl Soft Comput* 2011;11(8):5590–601. doi:10.1016/j.asoc.2011.04.005.
- [48] Deng X. Analyzing the monotonicity of belief interval based uncertainty measures in belief function theory. *Int J Intell Syst* 2018;33(9):1869–79.
- [49] Deng X, Jiang W. Dependence assessment in human reliability analysis using an evidential network approach extended by belief rules and uncertainty measures. *Ann Nucl Energy* 2018;117:183–93.
- [50] Jiang W. A correlation coefficient for belief functions. *Int J Approx Reason* 2018;103:94–106.
- [51] Zhang H, Deng Y. Engine fault diagnosis based on sensor data fusion considering information quality and evidence theory. *Adv Mech Eng* 2018;10(10). doi:10.1177/1687814018809184.
- [52] Pedrycz W, Bargiela A. Fuzzy fractal dimensions and fuzzy modeling. *Inf Sci* 2003;153:199–216. doi:10.1016/s0020-0255(03)00075-6.
- [53] Castillo O, Melin P. Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory. *IEEE Trans Neural Networks* 2002;13(6):1395–408. doi:10.1109/tnn.2002.804316.
- [54] Melin P, Castillo O. An intelligent hybrid approach for industrial quality control combining neural networks, fuzzy logic and fractal theory. *Inf Sci* 2007;177(7):1543–57. doi:10.1016/j.ins.2006.07.022.
- [55] Li Y, Deng Y. Generalized ordered propositions fusion based on belief entropy. *Int J Comput Commun Control* 2018;13(5):792–807.
- [56] Deng Y. Deng entropy. *Chaos Solitons Fractals* 2016;91:549–53.
- [57] Mo H, Deng Y. A new MADA methodology based on d numbers. *Int J Fuzzy Syst* 2018;20(8):2458–69.
- [58] Zhang H, Hu Y, Lan X, Mahadevan S, Deng Y. Fuzzy fractal dimension of complex networks. *Appl Soft Comput* 2014;25:514–18. doi:10.1016/j.asoc.2014.08.019.
- [59] Ivanovici M, Richard N. Fractal dimension of color fractal images. *IEEE Trans Image Process* 2011;20(1):227–35. doi:10.1109/tip.2010.2059032.
- [60] Bohdalova M, Gregus M. Fractal analysis of forward exchange rates. *Acta Polytechnica Hungarica* 2010;7(4):57–69.
- [61] Bogdan P, Jonckheere E, Schirmer S. Multi-fractal geometry of finite networks of spins: nonequilibrium dynamics beyond thermalization and many-body-localization. *Chaos Solitons Fractals* 2017;103:622–31. doi:10.1016/j.chaos.2017.07.008.
- [62] Song CM, Gallos LK, Havlin S, Makse HA. How to calculate the fractal dimension of a complex network: the box covering algorithm. *J Stat Mech-TheoryExp* 2007;16. doi:10.1088/1742-5468/2007/03/p03006.
- [63] Wen T, Jiang W. An information dimension of weighted complex networks. *Phys a-Stat MechAppl* 2018;501:388–99. doi:10.1016/j.physa.2018.02.067.
- [64] Lu L, Zhou T, Zhang QM, Stanley HE. The h-index of a network node and its relation to degree and coreness. *Nat Commun* 2016;7:7. doi:10.1038/ncomms10168.
- [65] Zhang W, Deng Y. Combining conflicting evidence using the DEMATEL method. *Soft Comput* vol., no. 2018. doi:10.1007/s00500-018-3455-8.
- [66] Zhou T, Liu J-G, Bai W-J, Chen G, Wang B-H. Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. *Phys Rev E* 2006;74(5). doi:10.1103/PhysRevE.056109.
- [67] Wei D, Chen X, Deng Y. Multifractality of weighted complex networks. *Chin J Phys* 2016;54(3):416–23. doi:10.1016/j.cjph.2016.05.004.
- [68] Xu S, Jiang W, Deng X, Shou Y. A modified physarum-inspired model for the user equilibrium traffic assignment problem. *Appl Math Model* 2018;55:340–53.
- [69] Yang H, Deng Y, Jones J. Network division method based on cellular growth and physarum-inspired network adaptation. *Int J Unconv Comput* 2018;13(6):477–91.
- [70] Ben-Naim E, Frauenfelder H, Toroczkai Z. *Complex networks*. Springer Berlin Heidelberg; 2004.
- [71] Silva FN, Costa LDF. Local dimension of complex networks. *Comput Sci* 2013.
- [72] Krackhardt D. Assessing the political landscape: structure, cognition, and power in organizations. *Adm Sci Q* 1990;35(2):342–69.
- [73] Kaiser M, Hilgetag CC. Spatial growth of real-world networks. *Phys Rev E* 2004;69(3):5. doi:10.1103/PhysRevE.69.036103.
- [74] Leslie PH. The treatment of ties in ranking problems. *Biometrika* 1945;33(3):239–51.
- [75] Lu L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T. Vital nodes identification in complex networks. *Phys Rep* 2016;650.
- [76] Hoeffding W. A non-parametric test of independence. *Ann Math Stat* 1948;19(4):546–57.
- [77] Hauskrecht M, Pelikan R, Valko M, Lyonsweiler J. Feature selection and dimensionality reduction in genomics and proteomics. Springer US; 2007.