

A novel method for identifying influential nodes in complex networks based on gravity model

Yuan Jiang(蒋沅), Song-Qing Yang(杨松青), Yu-Wei Yan(严玉为),Tian-Chi Tong(童天驰), and Ji-Yang Dai(代冀阳)
Citation: Chin. Phys. B, 2022, 31 (**5**): 058903. DOI: 10.1088/1674-1056/ac4226
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Detection of influential nodes with multi-scale information

Jing-En Wang(王静恩), San-Yang Liu(刘三阳), Ahmed Aljmiai, and Yi-Guang Bai(白艺光)
Chin. Phys. B, 2021, 30 (**8**): 088902. DOI: 10.1088/1674-1056/abff2d

LCH: A local clustering H-index centrality measure for identifying and ranking influential nodes in complex networks

Gui-Qiong Xu(徐桂琼), Lei Meng(孟蕾), Deng-Qin Tu(涂登琴), and Ping-Le Yang(杨平乐)
Chin. Phys. B, 2021, 30 (**8**): 088901. DOI: 10.1088/1674-1056/abea86

Shortest path of temporal networks: An information spreading-based approach

Yixin Ma(马一心), Xiaoyu Xue(薛潇雨), Meng Cai(蔡萌), and Wei Wang(王伟)
Chin. Phys. B, 2020, 29 (**12**): 128902. DOI: 10.1088/1674-1056/abb302

Influential nodes identification in complex networks based on global and local information

Yuan-Zhi Yang(杨远志), Min Hu(胡敏), Tai-Yu Huang(黄泰愚)
Chin. Phys. B, 2020, 29 (**8**): 088903. DOI: 10.1088/1674-1056/ab969f

Identifying influential spreaders in complex networks based on entropy weight method and gravity law

Xiao-Li Yan(闫小丽), Ya-Peng Cui(崔亚鹏), Shun-Jiang Ni(倪顺江)
Chin. Phys. B, 2020, 29 (**4**): 048902. DOI: 10.1088/1674-1056/ab77fe

---

A novel method for identifying influential nodes in complex networks based on gravity model

Yuan Jiang(蒋沅)[1], Song-Qing Yang(杨松青)[1,†], Yu-Wei Yan(严玉为)[1], Tian-Chi Tong(童天驰)[2], and Ji-Yang Dai(代冀阳)[1]

[1] *School of Information Engineering, Nanchang Hangkong University, Nanchang* 330063, *China*
[2] *School of Automation, Nanjing University of Technology, Nanjing* 210094, *China*

(Received 8 November 2021; revised manuscript received 7 December 2021; accepted manuscript online 11 December 2021)

How to identify influential nodes in complex networks is an essential issue in the study of network characteristics. A number of methods have been proposed to address this problem, but most of them focus on only one aspect. Based on the gravity model, a novel method is proposed for identifying influential nodes in terms of the local topology and the global location. This method comprehensively examines the structural hole characteristics and K-shell centrality of nodes, replaces the shortest distance with a probabilistically motivated effective distance, and fully considers the influence of nodes and their neighbors from the aspect of gravity. On eight real-world networks from different fields, the monotonicity index, susceptible-infected-recovered (SIR) model, and Kendall's tau coefficient are used as evaluation criteria to evaluate the performance of the proposed method compared with several existing methods. The experimental results show that the proposed method is more efficient and accurate in identifying the influence of nodes and can significantly discriminate the influence of different nodes.

Keywords: influential nodes, gravity model, structural hole, K-shell

PACS: 89.75.Hc, 89.75.Fb **DOI:** 10.1088/1674-1056/ac4226

1. Introduction

With the rapid development of network information technology, studies on complex networks has attracted attention in many fields.[1–4] As one of the core elements of a complex network, influential nodes have a significant impact on the structure and function of the network and are widely studied and applied, such as inhibiting disease spreading,[5] rumor controlling,[6] information dissemination,[7] and traffic network analysis.[8]

In different research backgrounds, scientists have proposed many measures to identify the influence of nodes. At present, degree centrality[9] is the commonly used measure, which is simple but not accurate. Betweenness centrality[10] and closeness centrality[11] are based on the number of shortest paths between nodes and the average shortest distance respectively, however, the information flow does not always transmit through the shortest path and its time complexity is high. Brockman *et al.*[12] proposed a probabilistically motivated effective distance to replace the shortest distance, which can effectively mine the dynamic interaction information between nodes. Kitsak *et al.*[13] proposed the K-shell decomposition method to determine the locations of nodes in the network, which has low time complexity, however its discrimination is too coarse. To improve the discrimination, Zeng *et al.*[14] proposed a mixed degree decomposition method, which decomposes K-shell based on the mixing degree of the remaining neighbors and the deleted neighbors in the network. Bae and Kim *et al.*[15] introduced the concept of neighborhood coreness in their improvement of the K-shell method, using the sum of K-shell values of nodes and their neighbors to identify the influence of nodes in the network. Wang *et al.*[16] proposed a vector centrality method based on the number of multi-order neighbor shells, using the form of vectors to represent the relative influence of nodes in a network. Burt *et al.*[17] proposed the theory of structural holes and designed the network constraint index to quantify the constraints that nodes are subjected to when they form structural holes, which can identify the bridging nodes in a network. Su *et al.*[18] proposed a method based on the structural holes of nodes and their neighbors, which considers both the degree centrality and topology of nodes and their neighbors and can effectively identify influential nodes in a network. Yang *et al.*[19] quantified the complexity of the network based on Tsallis entropy and combined the structural holes and global location information of nodes to identify influential nodes in the network.

Gravity model is derived from Newton's law of gravitation, which can represent the interaction between nodes in a complex network. If a node has more interaction with others, the node should be more influential. Based on gravity model, Li *et al.*[20] considered both neighborhood information and path information to identify influential nodes. Ma *et al.*[21] proposed a gravity model that regards the K-shell value of each node as its mass. Yan *et al.*[22] proposed a method to identify influential spreaders based on entropy weight method

†Corresponding author. E-mail: Young_SongQ@163.com

© 2022 Chinese Physical Society and IOP Publishing Ltd

and gravity law.

Based on the above research works, we propose a novel method for identifying influential nodes in complex networks based on gravity model, called EGMS. The EGMS method considers both the local topology and the global location of nodes in a network. In the EGMS method, the K-shell centrality is used to reflect the global location of nodes and the structure hole characteristics are used to represent the local topology of nodes. At the same time, to simplify the spatiotemporal complexity of a network and mine the implicit dynamic information in a network, the shortest distance is replaced by a probabilistically motivated effective distance. Then this method fully considers the influence of nodes and their neighbors from the aspect of gravity. By using the monotonicity index, susceptible-infected-recovered (SIR) model, and Kendall's tau coefficient as evaluation criteria, we compare the proposed method with seven existing identification methods on eight real-world networks. The experimental results show the excellent performance of the proposed method and the relationship between different methods.

The structure of this paper is as follows. In Section 2, we briefly introduce several existing identification methods and the concept of effective distance. In Section 3, a novel method for identifying influential nodes in complex networks based on gravity model is then proposed and presented. In Section 4, the details and results of experiments are then presented. Finally, conclusions will be given in Section 5.

2. Preliminaries

Given a complex network $G=(V,E)$, where V and E represent the set of nodes and edges in G respectively. n and m are the number of nodes and edges. The adjacency matrix of network G is denoted as $A_{n\times n}=(a_{ij})_{n\times n}$, where $a_{ij}=1$ if there is a connective relationship between node v_i and node v_j, and $a_{ij}=0$ otherwise.

2.1. Degree centrality

Degree centrality[9] (DC) is a simple measure for identifying influential nodes. The more DC a node possesses, the more influential it is. It is defined as follows:

$$DC(i)=\sum_{j=1}^{n}a_{ij}=k_i, \tag{1}$$

where k_i is the degree of node v_i. Hence node v_i's degree of next-nearest neighbors is $Q(i)=\sum_{\omega\in\Gamma(i)}k_\omega$, where $\Gamma(i)$ is the set of v_i's nearest neighbors.

2.2. Betweenness centrality

Betweenness centrality[10] (BC) focuses on the degree of clustering of the shortest path through a node. It is defined as follows:

$$BC(i)=\sum_{s\neq i\neq t}\frac{\sigma_{st}^{i}}{\sigma_{st}}, \tag{2}$$

where σ_{st} is the number of shortest paths between nodes v_s and v_t in the network, and σ_{st}^{i} is the number of those shortest paths that go through node v_i.

2.3. Closeness centrality

Closeness centrality[11] (CC) is defined as the reciprocal of the sum of the shortest distances between a node and all other nodes in the network. It is expressed as follows:

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

where d_{ij} denotes the shortest distance between nodes v_i and v_j.

2.4. Mixed degree decomposition method

Mixed degree decomposition (MDD) method was proposed by Zeng *et al.*,[14] it effectively improves the discrimination of the K-shell decomposition process by simultaneously considering the remaining degree value k_i^{r} and the removed degree value k_i^{e} of the node, which is defined as follows:

$$ks_i=k_i^{\mathrm{r}}+\lambda k_i^{\mathrm{e}}, \tag{4}$$

where λ is an adjustable parameter in range [0, 1], the MDD is consistent with the K-shell decomposition method if $\lambda=0$, and it is equivalent to degree centrality if $\lambda=1$. In this paper, we set the parameter $\lambda=0.7$ consistent with Ref. [11].

2.5. Structural holes theory

Structural hole theory was put forward by Burt in 1992 when he studied the competitive relationship in social networks.[17] In social networks, structural holes refer to non-redundant relationships between individuals, and there is a gap in complementary information. In order to quantify the control of the structural hole nodes on these relationships, Burt uses the network constraint index (NC) to measure the constraints formed by the nodes in the structural hole. The node is more likely to form structural holes if the index is smaller, and the influence of the node is greater. The network constraint index of node v_i is defined as follows:

$$RC_i=\sum_{j}\left(L_{ij}+\sum_{q(\neq i,j)}L_{iq}L_{qj}\right)^{2}, \tag{5}$$

where node v_q is the common neighbor node of node v_i and node v_j and L_{ij} denotes the weight ratio of node v_j to all of node v_i's neighbors. In order to make the influence of node v_i monotonically incremental to its network constraint coefficient, the network constraint index of node v_i can be further calculated as follows:

$$NC_i=1-RC_i. \tag{6}$$

When calculating the network constraint index of node v_i, considering the degree of the node and its neighbors, as well as the topological structure of the node, the weight ratio L_{ij} can be defined as follows:

$$L_{ij} = \frac{Q(j)}{\sum_{\upsilon \in \Gamma(i)} Q(\upsilon)}. \tag{7}$$

2.6. Gravity model

Gravity model[20] (GM) is inspired by Newton's law of gravity, and it identifies influential nodes by considering both degree centrality and the shortest distance, which is defined as follows:

$$GM(i) = \sum_{j \neq i} \frac{k_i k_j}{d_{ij}^2}. \tag{8}$$

2.7. Effective distance

Effective distance[12] (Edis) was proposed by Brockmann and Helbing in 2013 when studying the impact of complex networkdriven on the spreading process of epidemic diseases. Unlike conventional geographic distance and the shortest path distance, effective distance is a probabilistically motivated abstract distance. In the network where information flow is the interaction, it can effectively measure the mutual influence between nodes. The effective length $d_{j|i}$ from node v_i to a directly connected node v_j is defined as follows:

$$d_{j|i} = (1 - \log P_{j|i}) \geq 1, \tag{9}$$

where $P_{j|i}$ denotes the ratio of information flow from node v_i to node v_j, which can be calculated as follows:

$$P_{j|i} = \frac{a_{ij}}{k_i}, \quad (i \neq j), \tag{10}$$

where k_i denotes the degree of node v_i in the undirected network and the out-degree of node v_i in the directed network, a_{ij} is the element in the adjacency matrix $A_{n\times n}$. For indirectly connected node v_i and node v_j, the effective length between them will be transmitted along the shortest path, *e.g.*, $d_{j|i} = d_{k|i} + d_{j|k}$ like form.

The effective distance is defined by the shortest one of all effective lengths from node v_i to node v_j, which is expressed as follows:

$$D_{j|i} = \min_{\Gamma} \lambda(\Gamma), \tag{11}$$

where $\lambda(\Gamma)$ and Γ denote the set of all effective lengths and the sum of all the probably paths from node v_i to node v_j respectively. Usually, in the real-world networks, the $d_{j|i} \neq d_{i|j}$ and $D_{j|i} \neq D_{i|j}$.

3. Proposed method

3.1. The EGMS method

Based on the gravity model, a novel method is proposed for identifying influential nodes in terms of the local topology and the global location, called EGMS. The EGMS method comprehensively examines the structural hole characteristics and K-shell centrality of nodes, replaces the shortest distance with a probabilistically motivated effective distance, and fully considers the influence of nodes and their neighbors from the aspect of gravity. A flowchart of the proposed method is shown in Fig. 1, and the further details are as follows:

Step 1 Calculating the joint K-shell centrality of node pairs

K-shell centrality is a global measure describing the location of a node in a network, it can reflect both the relationship between each node and the whole network as well as the interaction between distinct nodes. In this paper, the K-shell centrality of each node is obtained by the MDD method mentioned in Subsection 2.4, and the joint K-shell centrality of node pairs can be obtained as follows:

$$\alpha(i,j) = \sqrt{ks_i + ks_j}. \tag{12}$$

Step 2 Calculating the joint network constraint index of node pairs

The network constraint index in structural hole theory is an identification metric based on local topological information of the network, it can be calculate by Eqs. (5)–(7), and the joint network constraint index between node v_i and node v_j can be calculated as follows:

$$\beta(i,j) = \sqrt{NC_i + NC_j}. \tag{13}$$

Step 3 Calculating the effective distance between all node pairs

In our proposed EGMS method, the shortest distance is replaced by the probabilistically motivated effective distance to simplify the spatiotemporal complexity of a network and mine the implicit dynamic information in a network. The effective distance from one node to another node can be obtained using Eqs. (9)–(11).

Step 4 Calculating the combined forces of nodes

By comprehensively considering the K-shell centrality, network constraints and effective distances, the interaction forces between node pairs can be obtained. Thus, the combined force of node v_i is defined as follows:

$$CG(i) = \sum_{j=1, j\neq i}^{N} \frac{\alpha(i,j) \times \beta(i,j)}{D_{j|i}^2}. \tag{14}$$

Step 5 Getting the final influence of nodes

Bae and Kim[15] introduced the concept of neighborhood coreness (Cnc+) in improving the K-shell method, which

measures the influence of nodes in a network by summing the K-shell values of their neighbors. In this paper, we borrow this concept and fully consider the influence of nodes' neighbors from the aspect of gravity. Based on the above, the final influence of node v_i can be calculated as follows:

$$EGMS(i) = \sum_{j\in\Gamma(i)} \sum_{j\in\Gamma(i)} CG(i). \tag{15}$$

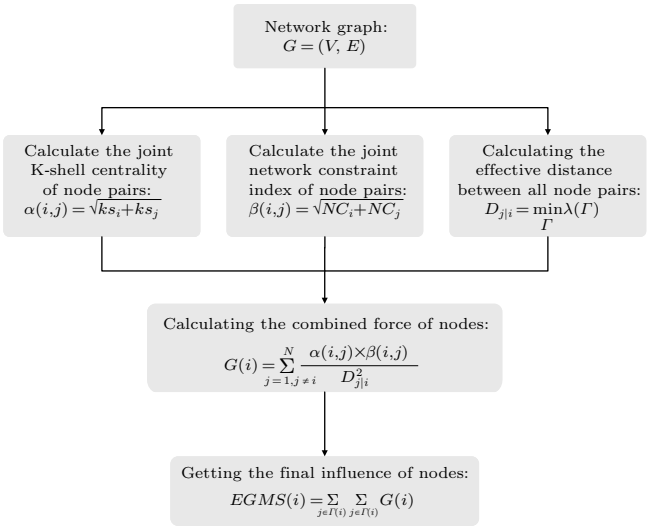


Fig. 1. The flow chart of our proposed EGMS method.

3.2. A simpe example

To better illustrate how the EGMS method works and to initially assess its validity and accuracy, a simple example network with 10 nodes and 14 edges is given in Fig. 2, and some properties of each node in the example network are shown in Table 1.

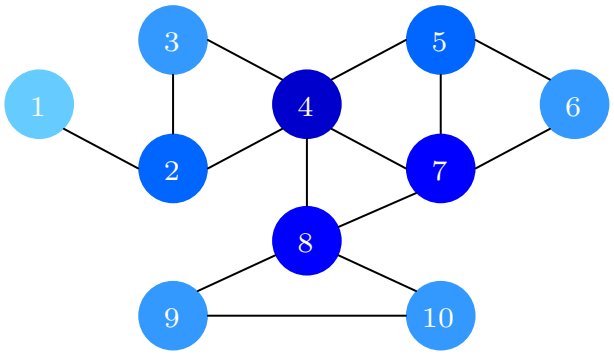


Fig. 2. A simple example network with 10 nodes and 14 edges. Nodes with larger degrees have darker colors.

Firstly, we calculate the joint K-shell centrality of node pairs by Eq. (12) and the joint network constraint index of node pairs by Eq. (13) respectively. The joint K-shell centrality and the joint network constraint index between node v_4 and node v_9 can be obtained as follows:

$$\begin{aligned}\alpha(4,9) &= \sqrt{ks_4+ks_9} = \sqrt{3.5+2} = 2.3452,\\ \beta(4,9) &= \sqrt{NC_4+NC_9} = \sqrt{0.4170+0.9815}\\ &= 1.1826.\end{aligned} \tag{16}$$

Secondly, by using Eqs. (9)–(11), the effective distance between node v_4 and node v_9 can be obtained as follows:

$$\begin{aligned}D_{4|9} &= \min\left\{D_{8|9}+D_{4|8}, D_{8|9}+D_{7|8}+D_{4|7},\ldots\right\}\\ &= \min\left\{1-\log\left(P_{8|9}\right)+1-\log\left(P_{4|8}\right), 1-\log\left(P_{8|9}\right)+1-\log\left(P_{7|8}\right)+1-\log\left(P_{4|7}\right)+,\ldots\right\}\\ &= 2-\log\left(P_{8|9}\right)-\log\left(P_{4|8}\right) = 4.0794,\\ D_{9|4} &= \min\left\{D_{8|4}+D_{9|8}, D_{7|4}+D_{8|7}+D_{9|8},\ldots\right\}\\ &= \min\left\{1-\log\left(P_{8|4}\right)+1-\log\left(P_{9|8}\right), 1-\log\left(P_{7|4}\right)+1-\log\left(P_{8|7}\right)+1-\log\left(P_{9|8}\right)+,\ldots\right\}\\ &= 2-\log\left(P_{8|4}\right)-\log\left(P_{9|8}\right) = 4.9957.\end{aligned} \tag{17}$$

The above results clearly show that $D_{4|9}$ and $D_{9|4}$ are not equal, as mentioned in Subsection 2.7. The results of the effective distance from node v_4 to the others are shown in Table 2.

Table 1. Properties of each node in the example network, including DC, MDD, and NC.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
DC	1	3	2	5	3	2	4	4	2	2
MDD	1	2.4	2	3.5	2.7	2	3.4	3.4	2	2
NC	0	0.2158	0.0665	0.5830	0.2690	0.1461	0.3988	0.4982	0.0185	0.0185

Table 2. Effective distance from node v_4 to the others in the example network.

	$D_{1\|4}$	$D_{2\|4}$	$D_{3\|4}$	$D_{4\|4}$	$D_{5\|4}$	$D_{6\|4}$	$D_{7\|4}$	$D_{8\|4}$	$D_{9\|4}$	$D_{10\|4}$
Edis	5.9069	3.3219	3.3219	0	3.3219	5.9069	3.3219	3.3219	6.3219	6.3219

Table 3. The combined forces of nodes in the example network.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
CG	0.2658	0.3352	0.3721	0.7188	0.5022	0.4042	0.6273	0.6402	0.2521	0.2521

Table 4. The final influence of nodes in the example network.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
EGMS	1.3567	3.8633	3.8337	2.2768	5.8719	4.0157	7.2073	6.5271	2.7428	2.7428

Then, the combined force of node v_4 can be obtained by Eq. (14) as follows:

$$CG(4) = \sum_{j=1, j\neq 4}^{10} \frac{\alpha(4,j) \times \beta(4,j)}{D_{j|4}^2} = 0.7188. \tag{18}$$

The combined forces of nodes are shown in Table 3.

Finally, as shown in Eq. (15), by accumulating the combined forces of node v_4's neighbors, the final influence of node v_4 can be obtained as follows:

$$EGMS(4) = \sum_{j\in\Gamma(4)} \sum_{j\in\Gamma(4)} CG(4) = 2.2768. \tag{19}$$

The final influence of nodes are shown in Table 4.

According to Table 1 and Table 4, the ranking lists of nodes in the example network measured by different methods are shown in Table 5. As shown in Table 5, the ranking lists of nodes measured by the EGMS method is is similar to other methods, indicating that the EGMS method can effectively identify the influence of nodes. Compared with DC and MDD, the influence of node v_7 and node v_8 is further discriminated in the EGMS method. And compared with NC, node v_2 and node v_6 are re-ranked in the EGMS method by considering the influence of their neighbors. Based on the above, the validity and accuracy of EGMS method is initially verified.

Table 5. The ranking lists of nodes in the example network measured by different methods.

Rank	1	2	3	4	5	6	7	8	9	10
DC	v_4	v_7	v_8	v_2	v_5	v_3	v_6	v_9	v_{10}	v_1
MDD	v_4	v_7	v_8	v_5	v_2	v_3	v_6	v_9	v_{10}	v_1
NC	v_4	v_8	v_7	v_5	v_2	v_6	v_3	v_9	v_{10}	v_1
EGMS	v_4	v_7	v_8	v_5	v_6	v_2	v_3	v_9	v_{10}	v_1

4. Experimental analysis

4.1. Dataset

In this section, eight real-world networks from different fields are chosen for experiments, including the Dolphins,[23] Polbooks,[24] Jazz,[25] USAir,[26] NetScience,[27] C.Elegans,[28] Email,[29] PowerGrid.[30] The statistical characteristics of eight real-world networks are shown in Table 6. The table includes information on the number of nodes n, the number of edges m, the average degree $\langle k\rangle$, the clustering coefficient c, the average shortest path length $\langle d\rangle$, the epidemic threshold β_{th} in SIR model, and the fixed infection probability β.

Table 6. Statistical characteristics of eight real-world networks.

Network	n	m	$\langle k\rangle$	c	$\langle d\rangle$	β_{th}	β
Dolphins	62	159	5.1290	0.2590	3.3570	0.1470	0.15
Polbooks	105	441	8.4000	0.4880	3.0788	0.0838	0.10
Jazz	198	2742	27.6970	0.6175	2.2350	0.0259	0.04
USAir	332	2126	12.8072	0.7494	2.7381	0.0225	0.03
NetScience	379	914	4.8232	0.7981	6.0419	0.1247	0.13
C.Elegans	453	2025	8.9404	0.6465	2.4553	0.0249	0.03
Email	1133	5451	9.6222	0.2540	3.6060	0.0535	0.05
PowerGrid	4941	6594	2.6691	0.0801	18.9892	0.2583	0.25

4.2. Comparison methods

To verify the validity and accuracy of our proposed EGMS method, seven methods are used as comparisons, including degree centrality (DC),[9] betweenness centrality (BC),[10] closeness centrality (CC),[11] mixed degree decomposition method (MDD),[14] network constraint index (NC),[18] neighborhood coreness (Cnc+),[15] and gravity model (GM).[20]

4.3. Evaluation criteria

To evaluate the performance of difference methods, several evaluation criteria are presented in this section, including monotonicity index,[15] susceptible–infected–recovered (SIR) model,[31] and Kendall's tau coefficient.[32]

The monotonicity index $M(R)$ is used to quantify the resolution of different identification methods, which is defined as follows:

$$M(R) = \left[1 - \frac{\sum_{r\in R} n_r(n_r-1)}{n(n-1)}\right]^2, \tag{20}$$

where R is the influence ranking vector of each node obtained by the influence identification method, n is the total number of nodes, and n_r is the number of nodes with the same influence. The value of $M(R)$ in range $[0,1]$, which will be higher if the identification method can better distinguish the influence of all nodes in the network.

The SIR model[31] is used to examine the spreading ability of nodes. In this model, each node in the network can be in one of the following three states: susceptible, infected, or recovered. In the initial stage of spreading, only one infected node in the network, and all other nodes are susceptible. In each timestamp, each one infected node to infect its susceptible neighbors with the infection probability β. At the same time, the infected node recovers with probability μ. This process is repeated until there are no susceptible nodes in the network. At the end of SIR spreading process, the number of recovered nodes is regarded as the spreading ability of nodes. For simplicity, the value of the recovery probability is set to $\mu = 1$. The spreading process cannot be carried out normally if the infection probability β is too small or too large, do the spreading ability of each node cannot be measured accurately. Therefore, we set the epidemic threshold is $\beta_{\text{th}} \approx \langle k \rangle / \langle k^2 \rangle$, where $\langle k \rangle$ and $\langle k^2 \rangle$ denote the average degree and average second-order degree of the nodes respectively. The value of β is set near β_{th}. As show in Table 6, the β_{th} and β for different networks are given.

In order to evaluate the accuracy of the identification method, Kendall's tau coefficient[32] is used to measure the correlation between the ranking list of nodes' influence obtained by each identification method and the ranking list of nodes' spreading ability obtained from the SIR model. Suppose there are two sequences X and Y containing n nodes, where $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$, and the elements in sequence X and sequence Y are constructed into a new sequence $XY = ((x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n))$. For any pair of elements $XY_i = (x_i, y_i)$ and $XY_j = (x_j, y_j)$ in the sequence XY, if ($x_i > x_j$ and $x_i > y_j$) or ($x_i < x_j$ and $x_i < y_j$), the pair is considered to be concordant. On the contrary, if ($x_i > x_j$ and $x_i < y_j$) or ($x_i < x_j$ and $x_i > y_j$), this pair is considered to be discordant. If ($x_i = x_j$ and $x_i = y_j$), it is considered that the pair is neither concordant nor discordant. Kendall's tau coefficient τ is defined as

$$\tau(X, Y) = \frac{n_{\text{c}} - n_{\text{d}}}{n(n-1)/2}, \tag{21}$$

where n_{c} and n_{d} are the numbers of concordant pairs and discordant pairs respectively. The value of τ in range $[-1, 1]$. The two sequences are completely positively correlated if $\tau = 1$, that is to say, the closer the value of τ is to 1, the higher the accuracy. On the contrary, the two sequences are completely negatively correlated.

4.4. Experiments

4.4.1. Examining the resolution of different identification methods

In this experiment, the monotonicity index[13] M is used to examine the resolution of the seven identification methods and our proposed EGMS method, *i.e.*, the ability to distinguish the influence of nodes. Table 7 records the resolution of the different methods in eight real-world networks, and figure 3 shows the difference more intuitively. As shown in Fig. 3, the EGMS method performs very well in the eight networks. In addition, the resolution of CC and GM is also good. Compared with others, the EGMS method gives a higher M value in most cases, and the value of M (EGMS) is very close to or even equal to 1 in all cases. Therefore, the EGMS method performs excellent in distinguish the influence of nodes.

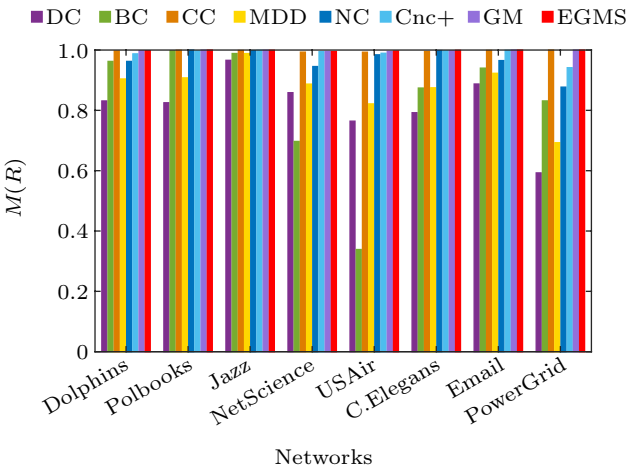


Fig. 3. The monotonicity index M of eight identification methods in eight real-world networks.

Table 7. The monotonicity index M of different identification methods.

Network	M (DC)	M (BC)	M (CC)	M (MDD)	M (NC)	M (Cnc+)	M (GM)	M (EGMS)
Dolphins	0.8312	0.9623	0.9968	0.9041	0.9623	0.9873	0.9979	0.9979
Polbooks	0.8252	0.9974	0.9993	0.9077	1	0.9971	0.9996	1
Jazz	0.9659	0.9885	0.9986	0.9883	0.9983	0.9993	0.9995	0.9994
USAir	0.8586	0.6970	0.9931	0.8871	0.9453	0.9945	0.9951	0.9951
NetScience	0.7642	0.3387	0.9928	0.8215	0.9841	0.9893	0.9956	0.9967
C.Elegans	0.7922	0.8740	0.9952	0.8748	0.9983	0.9975	0.9977	0.9991
Email	0.8874	0.9400	0.9988	0.9229	0.9650	0.9991	0.9999	0.9999
PowerGrid	0.5927	0.8313	0.9999	0.6928	0.8770	0.9414	0.9999	0.9999

4.4.2. The comparisons of different identification methods and SIR model

In this experiment, the SIR model is used to obtain the spreading ability of nodes and Kendall's tau coefficient is used to quantify the correlation between the spreading ability and influence of nodes in the networks. The stronger the correlation, the more accurate the identification method is. Due to the randomness of the spreading process of the SIR model, this experiment repeats the spreading process for each node and then takes the average value. For networks $|V| < 100$, the independent experiment is executed 10^4 times; for $|V| < 10^3$, the independent experiment is executed 10^3 times. Table 8 shows Kendall's tau coefficient between different identification methods and the SIR model with the fixed infection probability in eight real-world networks, where σ denotes the ranking lists of the spreading ability of nodes in the networks. The values of the fixed infection probability are shown in Table 6.

In order to further verify the accuracy and reliability of the proposed method, a series of Kendall's tau coefficients between different identification methods and SIR model can be obtained by changing the infection probability β. The results are shown in Fig. 4, where the black dashed line denotes the epidemic threshold β_{th}. From which we can see, the EGMS method outperforms all the methods when the infection probability β takes values near the epidemic threshold β_{th}. This indicates that the influence of nodes identified by the EGMS method has a significant correlation with the spreading ability of nodes. When the infection probability is far greater than or far less than the epidemic threshold, DC, MDD and GM show better correlation than EGMS in six networks except USAir and PowerGrid. In addition, Cnc+ performed slightly worse than EGMS in most networks and BC performed worst in all cases. According to the above analysis, the EGMS method can identify the influence of nodes more accurately and reliably.

Table 8. Kendall's tau coefficient between different identification methods and SIR model with the fixed infection probability.

Network	τ (DC, σ)	τ (BC, σ)	τ (CC, σ)	τ (MDD, σ)	τ (NC, σ)	τ (Cnc+, σ)	τ (GM, σ)	τ (EGMS, σ)
Dolphins	0.7768	0.5643	0.8117	0.8181	0.7282	0.8611	0.8985	0.9561
Polbooks	0.7528	0.3579	0.6502	0.8029	0.7037	0.9098	0.8286	0.9477
Jazz	0.8185	0.4628	0.7950	0.8503	0.8216	0.9132	0.8643	0.9568
USAir	0.6982	0.4744	0.8414	0.7175	0.7929	0.9047	0.8250	0.9454
NetScience	0.5931	0.2844	0.4997	0.6141	0.6580	0.5943	0.7869	0.8905
C.Elegans	0.6376	0.4711	0.6933	0.6521	0.6251	0.8127	0.7944	0.8667
Email	0.8000	0.6467	0.8315	0.8115	0.7735	0.8980	0.8387	0.9313
PowerGrid	0.5469	0.4167	0.3916	0.5786	0.4298	0.7341	0.6950	0.8255

4.4.3. Kendall's τ coefficient for different identification methods to varying scales of topranking nodes

In this experiment, the scale of nodes examined by Kendall's tau coefficient is adjusted, which is changed from whole nodes to partial nodes. The correlation results between the top nodes in different scales obtained by different identification methods and the ranking list of the spreading ability of nodes are further studied. The scale of topranking nodes is denotes as L, which is set to 0.05–1. Figure 5 shows the values of Kendall's tau coefficient for different identification methods to varying scales of topranking nodes nodes. As shown in Fig. 5, the EGMS method performs well at various scales of topranking nodes and performs better than all other methods at higher values of L.

4.4.4. The relationship between EGMS and other identification methods

In the final experiment, the relationship between the EGMS method and other identification methods is analyzed. The results are shown in Fig. 6, where each point denotes one node in the network, the color of each point denotes the spreading ability derived by the SIR model, and the horizontal and vertical axes are denote the values of nodes' influence in the network obtained by the EGMS method and other methods respectively. The EGMS method has a strong correlation with one method if the nodes in the figure are linearly distributed. As shown in the distribution of the nodes in Fig. 6, strong correlations are found between EGMS and DC, CC, MMD, and NC, respectively. This indicates that the EGMS method not only takes into account the degree centrality, location information and the closeness between nodes sufficiently, but also has the ability to identify bridging nodes in the network. Moreover, the apparent linear relationship between EGMS and Cnc+ and GM indicates that the node rankings are overall similar, and it also demonstrates that EGMS can identify the influence of nodes effectively. From Fig. 6, it can also find that some nodes have large values of BC, but their spreading ability is not very strong, so it is not accurate enough to identify the influence of nodes by BC alone. Therefore, the EGMS method is more advantageous than other methods in identifying the influence of nodes in the network.

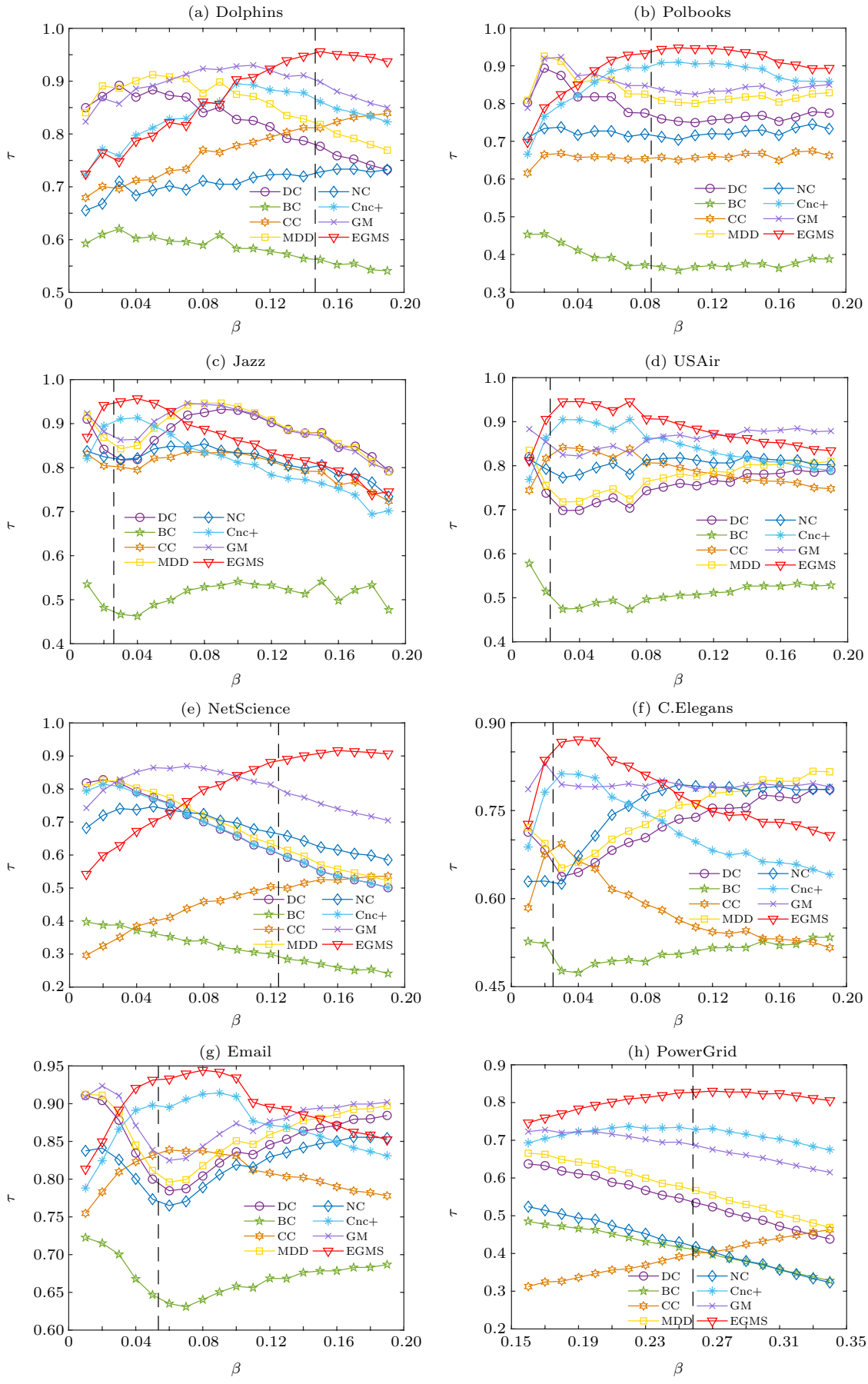


Fig. 4. Kendall's τ coefficient between different identification methods and SIR model with different infection probabilities: (a) Dolphins; (b) Polbooks; (c) Jazz; (d) USAir; (e) NetScience; (f) C.Elegans; (g) Email; (h) PowerGrid.

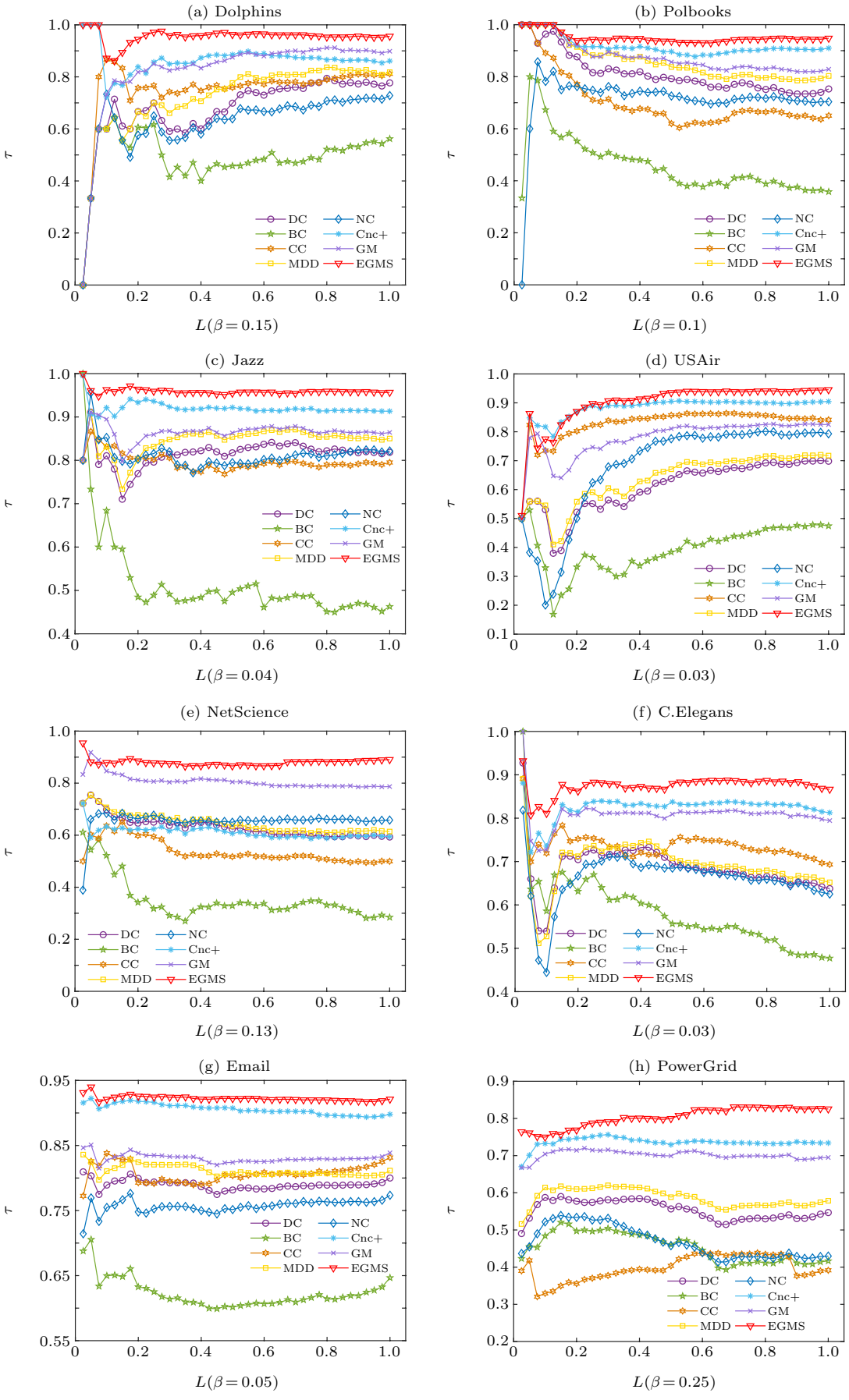


Fig. 5. Kendall's τ coefficient for different identification methods to varying scales of top-ranking nodes: (a) Dolphins; (b) Polbooks; (c) Jazz; (d) USAir; (e) NetScience; (f) C.Elegans; (g) Email; (h) PowerGrid.

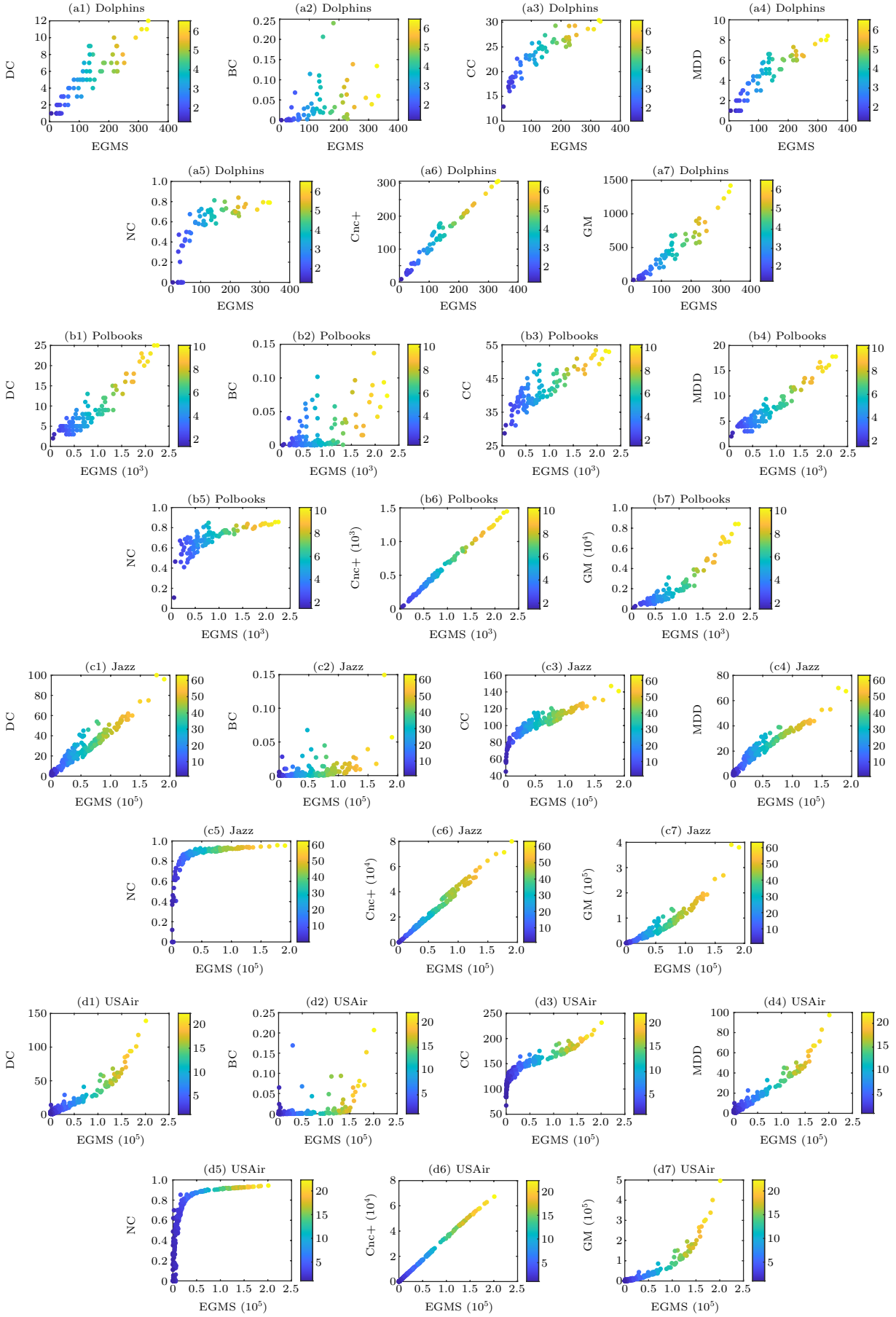


Fig. 6. The relationship between EGMS and other identification methods: (a1)–(a7) Dolphins, $\beta = 0.15$; (b1)–(b7) Polbooks, $\beta = 0.1$; (c1)–(c7) Jazz, $\beta = 0.04$; (d1)–(d7) USAir, $\beta = 0.03$.

5. Conclusions and discussion

In this paper, we proposed a novel method for identifying influential nodes in complex networks based on gravity model, called EGMS. For the shortcoming that the existing methods concentrate only on only one aspect, the EGMS method considers both the local topology and the global location information, comprehensively examines the structural hole characteristics and K-shell centrality of nodes At the same time, the shortest distance is replaced by a probabilistically motivated effective distance to simplify the spatiotemporal complexity of a network and mine the implicit dynamic information in a network. Then the EGMS method fully considers the influence of nodes and their neighbors from the aspect of gravity to optimize the utilization of available resources. In addition, the EGMS method also has the ability to identify bridging nodes in the network. From the experimental results, the EGMS method can better distinguish differences in the influence of nodes than other identification methods. Further, the EGMS method performs excellent in identifying influential nodes at various scales of topranking nodes. Finding a more effective method combining network topology and propagation dynamics to identify influential nodes in the network is still an open issue. How to improve our proposed EGMS method to reduce its time complexity is our future work.

Acknowledgment

Project supported by the National Natural Science Foundation of China (Grant Nos. 61663030 and 61663032).

References

[1] Liu Y Y, Slotine J J and Barabasi A L 2011 *Nature* **473** 167
[2] Gao J, Zhang Y C and Zhou T 2019 *Phys. Rep.* **817** 1
[3] Vermeulen R, Schymanski E L, Barabasi A L and Miller G W 2020 *Science* **367** 392
[4] Liu S T and Wang P 2018 *Fractal Control Theory* (Singapore: Springer Nature) pp. 9–24
[5] Shi H J, Duan Z S, Chen G R and Li R 2009 *Chin. Phys. B* **18** 03309
[6] Hosni E I A, Li K and Ahmad S 2020 *Inform. Sci.* **512** 1458
[7] Liu S T, Zhang Y P and Liu C 2020 *Fractal Control and Its Applications* (Singapore: Springer Nature) pp. 129–163
[8] Yan Y W, Jiang Y, Yu R B, Yang S Q and Hong C 2021 *Chin. Phys. B* **31** 018901
[9] Bonacich P 1972 *J. Math. Sociology* **2** 113
[10] Freeman L C 1977 *Sociometry* **40** 35
[11] Opsahl T, Agneessens F and Skvoretz J 2010 *Social Networks* **32** 245
[12] Brockmann D and Helbing D 2013 *Science* **342** 1337
[13] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E and Makse H A 2010 *Nat. Phys.* **6** 888
[14] Zeng A and Zhang C J 2013 *Phys. Lett. A* **377** 1031
[15] Bae J and Kim S 2014 *Physica A* **395** 549
[16] Wang K L, Wu C X, Ai J and Su Z 2019 *Acta Phys. Sin.* **68** 196402 (in Chinese)
[17] Burt R S, Kilduff M and Tasselli S. 2013 *Ann. Rev. Psychol.* **64** 527
[18] Su X P and Song R R 2015 *Acta Phys. Sin.* **64** 020101 (in Chinese)
[19] Yang S Q, Jiang Y, Tang T C, Yan Y W and Gan G S 2021 *Acta Phys. Sin.* **70** 216401 (in Chinese)
[20] Li Z, Ren T, Ma X Q, Liu S M, Zhang Y X and Zhou T 2019 *Sci. Rep.* **9** 8387
[21] Ma L L, Ma C, Zhang H F and Wang B H 2016 *Physica A* **451** 205
[22] Yan X L, Cui Y P and Ni S J 2020 *Chin. Phys. B* **29** 048902
[23] Lusseau D, Schneider K, Boisseau O, Haase P, Slooten E and Dawson S 2003 *Behav. Ecol. Sociobiol.* **54** 396
[24] Girvan M and Newman M E J 2002 *Proc. Nati. Acad. Sci.* **99** 7821
[25] Gleiser P M and Danon L 2003 *Complex Syst.* **06** 565
[26] Colizza V, Pastor-Satorras R and Vespignani A 2007 *Nat. Phys.* **3** 276
[27] Newman M E J 2006 *Phys. Rev. E* **64** 036104
[28] Duch J and Arenas A 2005 *Phys. Rev. E* **72** 027104
[29] Guimera R, Danon L, Diaz-Guilera A, Giralt F and Arenas A 2003 *Phys. Rev. E* **68** 065103
[30] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[31] Pastor-Satorras R and Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
[32] Knight W R 1966 *J. Amer. Statist. Associat.* **61** 436