



# A mechanics model based on information entropy for identifying influencers in complex networks

Shuyu Li<sup>1,2</sup> · Fuyuan Xiao<sup>2</sup>

Accepted: 5 January 2023 / Published online: 30 January 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

The network, with some or all characteristics of scale-free, self-similarity, self-organization, attractor and small world, is defined as a complex network. The identification of significant spreaders is an indispensable research direction in complex networks, which aims to discover nodes that play a crucial role in the structure and function of the network. Since influencers are essential for studying the security of the network and controlling the propagation process of the network, their assessment methods are of great significance and practical value to solve many problems. However, how to effectively combine global information with local information is still an open problem. To solve this problem, the generalized mechanics model is further improved in this paper. A generalized mechanics model based on information entropy is proposed to discover crucial spreaders in complex networks. The influence of each neighbor node on local information is quantified by information entropy, and the interaction between each node on global information is considered by calculating the shortest distance. Extensive tests on eleven real networks indicate the proposed approach is much faster and more precise than traditional ways and state-of-the-art benchmarks. At the same time, it is effective to use our approach to identify influencers in complex networks.

**Keywords** Complex network · Crucial spreaders identification · Gravity model · Information entropy

## 1 Introduction

A complex network involves multidisciplinary knowledge and theoretical foundations, which can effectively model different components or factors in a complex system [1]. There is no denying that complex networks have been applied diffusely in dissimilar fields such as natural sciences [2–4], social sciences [5, 6] and medical sciences [7–9]. With the continuous rise of research on complex networks, influencers identification, as an indispensable branch of complex networks, has become increasingly important [10, 11]. Not only does the identification of key nodes have high guiding value in theory, but also it has comprehensive applications in diverse fields such as the transmission of infectious disease [12], cyber security [13], biological

information [14], and social networking services [15–17]. Therefore, a large number of excellent algorithms have been gradually proposed. These algorithms are approximately classified into neighborhood-based centrality [18, 19], path-based centrality, and eigenvectors. The advantages and specific steps of each algorithm are described in Section 2.

Recently, some interaction-based gravity model algorithms have been put forward. The gravitational force of one node on another is proportional to the degree and inversely proportional to the distance between them in the gravity model. The algorithm based on the law of universal gravitation were put forward by Li et al. [20]. In addition, Liu et al. [21] further improved it. The high computational complexity is weakened, and the noise that may be generated by long-distance interactions is reduced. Li et al. [22] put forth a generalized gravity model (GGC for short), which takes into account the local clustering coefficient as the local information. But the parameters are more difficult to determine.

However, most of the above methods only consider local information or global information, which has certain limitations. The method that only considers local information creates the advantage of easy calculation. But it does not

✉ Fuyuan Xiao  
xiaofuyuan@cqu.edu.cn; doctorxiaofy@hotmail.com

<sup>1</sup> Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai, 200092, China

<sup>2</sup> School of Big Data and Software Engineering, Chongqing University, Chongqing, 401331, China

measure the structural information of the entire network, which leads to partial and inaccurate results. In addition, in some applications like the spread of infectious diseases, it is necessary to consider the global impact of nodes. This process needs to calculate how many other nodes are infected by infectious nodes within a certain period of time, which takes into account global information. At the same time, it is not feasible to consider only the global impact, which leads to high complexity. In some cases, this limitation will hinder its application. For example, it is not suitable for large-scale networks. Therefore, there is an open issue that is how to take advantage of local and global conditions to evaluate the influencers in complex networks. In order to settle the problem, a generalized mechanics model based on information entropy is proposed (named as Information Entropy-based Gravity Model, INEG for short), which combines local and global information. INEG not only makes the obtained results more accurate, but also reduces the computational complexity.

The main contribution of the paper is that information entropy is used to measure the uncertain information around the nodes. Uncertain information can be regarded as the uncertainty of the node spreading information to its neighbors. In other words, it is the probability that the neighbor is selected by the node. In the local information, INEG takes into consideration the influence of neighbor nodes in detail. The relative score is introduced in the global information, which is expressed as the weight of each node in the entire network. All in all, our main innovation is an effective combination of local and global information. What's more, another highlight of this paper is that location information is introduced. The degree of a node is treated as mass in the traditional gravity model, which is rough. In fact, the influence of a node is not only related to its degree, but also to its position in the network. Nodes at the core of the network tend to be more important than edge nodes because of the higher connection density between core nodes. Further, nodes with higher connection density are more able to influence other nodes. For example, there are two nodes of the same degree at different locations in the network. Obviously, a node at the core of the network has more influence than a node at the edge of the network. Therefore, location information needs to be considered in identifying critical nodes. Through the proposed inheritance rule, the obtained  $b$ -index is used to record the distance between the node and the peripheral part. The  $f$ -index is used to correct nodes with low  $b$ -index but close to the core of the network. Finally, the node mass with location information is generated according to the mixed index composed of  $b$  and  $f$ .

The rest of this article is organized as follows. First of all, Section 2 describes some preliminary knowledge and basic algorithms of complex networks briefly. At the same time,

three state-of-the-art approaches are given. Furthermore, the improved INEG method is presented in Section 3. Then, Section 4 demonstrates the rationality of our approach through several experiments. Some applications of this method are discussed in Section 5. Finally, the conclusions and future prospects are analysed in Section 6.

## 2 Related work

The representative methods based on neighborhood centrality include degree centrality (DC for short) [23], K-shell decomposition [24–28], and H-index [29] methods. DC holds the view that a node with more neighbors has a greater influence [23]. Nevertheless, the main disadvantage of DC is that only local information is considered. The K-shell decomposition [24] cogitates the location information of the nodes in the network. It claims that a node with a more central position in the network generally has a greater influence. Notwithstanding, its degree is smaller. But its result is too rough to discriminate nodes. The H-index deems that nodes connected to multiple neighbors with high degrees are more influential. Besides, closeness centrality (CC for short) [30] and betweenness centrality (BC for short) [31] are two representative path-based centralities ways. CC [30], which takes into account the global information of the network, evaluates the significance of one node by summarizing all the distances between it and the other nodes in the network. While BC [31] cogitates the number of shortest paths passing through a target node. Two well-studied methods based on eigenvector are the eigenvector centrality (EC for short) [32] method and the PageRank algorithm (PC for short) [33, 34]. For EC [32], the influence of a node rests with both the number of its neighbor nodes and their significances. But the information of other nodes connected to a node is completely used to evaluate the importance of the node. PC [33] ranks web pages on the basis of the link relationships of web pages. For PC, the importance of a page in the World Wide Web is determined by the number and quality of the pages that link to it. Hence, if a page is pointed to by many high-quality pages, the quality of this page is also high.

In the following subsections, firstly, the specific steps of the above methods are described. Then the recent algorithms concerning generalized mechanics model are reviewed. Lastly, two state-of-the-art algorithms are introduced.

### 2.1 The classic measures

#### 2.1.1 Degree centrality

DC, which describes the direct influence of nodes, is the simplest and most intuitive way to evaluate the importance of nodes in the network. It claims that a node with a higher

degree can directly affect more neighbors, which means it is more crucial. In order to compare nodes that have the same degree in networks of different scales, a normalized DC [23] is given, which is defined as follows:

$$DC_i = \frac{k_i}{n-1} \quad (1)$$

where  $k_i = \sum_j a_{ij}$  is the degree of node  $i$ ,  $a_{ij}$  stands for the connection between node  $i$  and node  $j$ , and the denominator  $n-1$  is the maximum degree value possible for the node.

### 2.1.2 K-shell algorithm

The K-shell algorithm emphasizes the location information of nodes and quantifies the importance of nodes by assigning them to different shells. The method considers that the nodes in the center of the network are more important, while the nodes closer to the edge are less important. Nodes with larger shells are closer to the network core. The specific steps of the K-shell algorithm are as follows. In the first step, nodes with degree  $k=1$  and their connected edges are sequentially removed from the network. Since the first step in the removal process may reduce the degree of the node, it is necessary to continue to remove the node with degree  $k \leq 1$  and its connected edges. In the second step, when there are no nodes with degree  $k \leq 1$  in the network, create  $1-shell$  for the nodes removed in the first step and let their  $shell=1$ . In the third step, the degree value is incremented by 1, and the above process is repeated until each node has a unique corresponding shell. After the execution of the above process,  $2-shell$ ,  $3-shell$  and so on can be obtained. Finally, all nodes in the network are divided into different shells, and the nodes in each shell have the same shell value. Obviously, nodes in the same shell are considered to have the same importance, while nodes in the inner shell have higher influence.

### 2.1.3 Closeness centrality

CC determines the significance of a node by summarizing the distance between the target node and all other nodes. If the average distance between a node and other nodes is smaller, the influence on the network is greater. Given a connected network with  $n$  nodes, the CC [30] of node  $i$  is defined as:

$$CC_i = \frac{n-1}{\sum_{j \neq i} d_{ij}} \quad (2)$$

where  $d_i$  stands for the average shortest distance from node  $i$  to other nodes in the network, and  $d_{ij}$  is the shortest path distance between node  $i$  and node  $j$ .

### 2.1.4 Betweenness centrality

BC describes the flow transmitted through the shortest path between pairs of nodes in the network. It evaluates the importance of nodes by calculating the concentration of paths. BC clings to the perspective that a node would have a greater influence if there are more shortest paths pass through it. The BC [31] of node  $i$  is defined as:

$$BC_i = \sum_{i \neq s, i \neq t, s \neq t} \frac{g_{st}}{G_{st}} \quad (3)$$

where  $G_{st}$  stands for paths between node  $s$  and node  $t$ , and  $g_{st}$  is the shortest paths between  $s$  and  $t$  via node  $i$ .

### 2.1.5 Eigenvector centrality

The EC focuses on the surrounding environment where the node is located, such as the quantity and quality of neighbor nodes. And its score value is related to the sum of the scores of its neighbors. Given the adjacency matrix  $A$  of a network, the EC [32] of node  $i$  is defined as follows:

$$EC_i = x_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j \quad (4)$$

where  $x_i$  is recorded as the importance metric value of node  $i$ , which is the value of the  $i$ th term in the normalized eigenvector of the largest eigenvalue of  $A$ .  $\lambda$  is a non-zero constant, which meets the following requirements when it reaches a steady state after multiple iterations:

$$Ax = \lambda x \quad (5)$$

where  $x$  is the eigenvector corresponding to the eigenvalue  $\lambda$  of matrix  $A$ .

### 2.1.6 PageRank algorithm

PC is the most famous algorithm in the field of web page ranking. At the initial moment, each node is given the same PR, and then iterated. At each step, the current PR of each node is equally assigned to all the nodes it points to. The process will end until the PR of each node is stable. The new PR of each node is the sum of its obtained PR, so the definition of the PR of node  $i$  at time  $t$  is as follows [33]:

$$PR_i(t) = \sum_{j=1}^n a_{ji} \frac{PR_j(t-1)}{k_j^{out}} \quad (6)$$

where  $k_j$  stands for the out degree of node  $j$ .

## 2.2 The measures based on gravity model

### 2.2.1 Gravity centrality algorithm

Inspired by the law of universal gravitation, gravity centrality (GC for short) regards the node’s degree as mass, and the shortest path distance in the network as the distance between two nodes. In order to solve the time-consuming and noise problem, a truncation radius is introduced here. The GC of node  $i$  is defined as [20]:

$$GC_i = \sum_{i \neq j, d_{ij} \leq R} \frac{k_i \times k_j}{d_{ij}^2} \tag{7}$$

where  $k_i$  and  $k_j$  denote the degree of node  $i$  and node  $j$  respectively,  $d_{ij}$  represents the shortest path distance between node  $i$  and node  $j$ , and  $R$  stands for the affected radius. Li et al. [20] found an empirical relationship between the truncation radius  $R$  and the average path length  $\langle d \rangle$  through a large number of experiments. Therefore, the best results are obtained when  $R$  is set to  $0.5 \langle d \rangle$ .

### 2.2.2 Weighted gravity centrality algorithm

The weighted gravity centrality (WGC for short) assigns a certain weight to each node through the eigenvector, which is an improvement to GC, and it can be defined as follows [21]:

$$WGC_i = \sum_{i \neq j, d_{ij} \leq 0.5 \langle d \rangle} e_i \times \frac{k_i \times k_j}{d_{ij}^2} \tag{8}$$

where  $e_i$  is the  $i$ th value of the normalized eigenvector corresponding to the largest eigenvalue of matrix  $A$ .

### 2.2.3 Generalized gravity centrality algorithm

GGC [22] regards the propagation ability of the node as the mass. A network is given, and it is defined as:

$$GGC_i = \sum_{i \neq j, d_{ij} \leq 0.5 \langle d \rangle} \frac{e^{-\alpha C_i} \times k_i \times e^{-\alpha C_j} \times k_j}{d_{ij}^2} \tag{9}$$

where  $d_{ij}$  indicates the distance between node  $i$  and node  $j$ , and  $R$  is the influenced radius.  $\alpha$  ( $\alpha \geq 0$ ) is a parameter, which can be flexibly modified in practical applications.  $C_i$  and  $C_j$  are the clustering coefficients, where  $C_i = \frac{2n_i}{k_i(k_i-1)}$ ,  $C_j = \frac{2n_j}{k_j(k_j-1)}$ ,  $k_i$  and  $k_j$  denote the degree of node  $i$  and node  $j$  respectively.

## 2.3 Two state-of-the-art measures

### 2.3.1 Local information dimensionality

The local information dimensionality (LID for short) [35] uses Shannon entropy [36, 37] to consider the number of nodes in each box. The information [38–42] of the nodes is adequately measured [43]. The LID of node  $i$  is defined as:

$$LID_i = -\frac{d}{d \ln l} \left( -\frac{n_i(l)}{N} \ln \frac{n_i(l)}{N} \right) \tag{10}$$

where  $d$  represents the derivative,  $l$  stands for the size of the box,  $n_i(l)$  denotes the number of nodes in the box, and  $N$  shows the total amount of nodes in the network.

### 2.3.2 Fuzzy local dimension

Because the distance from each node to the central node is diverse, the contribution of each node is also different. As a result, fuzzy sets [44–46] are used for assigning nodes that contribute to the local dimensionality [47] in the fuzzy local dimension (FLD for short) [48]. The definition of FLD is as follows:

$$FLD_i = \frac{d}{d \log r_t} \log \frac{\sum_{j=1}^N A_{ij}(\varepsilon)}{N_{i,r}} \tag{11}$$

where  $r_t$  stands for the radius from the center node  $i$ ,  $\varepsilon$  shows the size of the box, and  $N_{i,r}$  represents the number of real nodes when the shortest distance between  $i$  and  $j$  is less than the  $\varepsilon$ .  $A_{ij}(\varepsilon)$  represents the membership function when the distance from  $j$  to  $i$  is less than  $\varepsilon$ , where  $A_{ij}(\varepsilon) = \exp(-\frac{d_{ij}^2}{\varepsilon^2})$ .

## 3 Proposed method

### 3.1 Algorithmic procedure

In this paper, a generalized mechanics model (named as Information Entropy-based Gravity Model) based on information entropy is proposed. The highlight of our approach is that the influence of each neighbor node is considered in the local information. INEG uses entropy to score each neighbor node, and finally each node is assigned to an objective score. The global information is quantified as the distance between each pair of nodes, and the objective score of each node is converted into a relative score to modify the relative interaction between nodes. What’s more, location information is introduced to form the mass. The major contribution of this model is to combine local information and global information effectively, and form the relative score of each node objectively. It consists

of the following five steps, and the flowchart is shown in Fig. 1.

Step 1: Construct the network.

Input a real connected undirected network, and output the adjacency matrix of this network.

Step 2: Calculate the degree distribution and the shortest distance.

The adjacency matrix is input, the degree of each node and the shortest distance between different nodes are output.

Step 3: Calculate the information entropy and get the weight.

The adjacency matrix and the degree distribution matrix as input, then the information entropy of each node is calculated. Finally, the weight of each node is output.

Step 4: Calculate the mass of the nodes.

Input the degree distribution matrix from the second step and output the mass of nodes with position information.

Step 5: Calculate the interaction force and get the influence node.

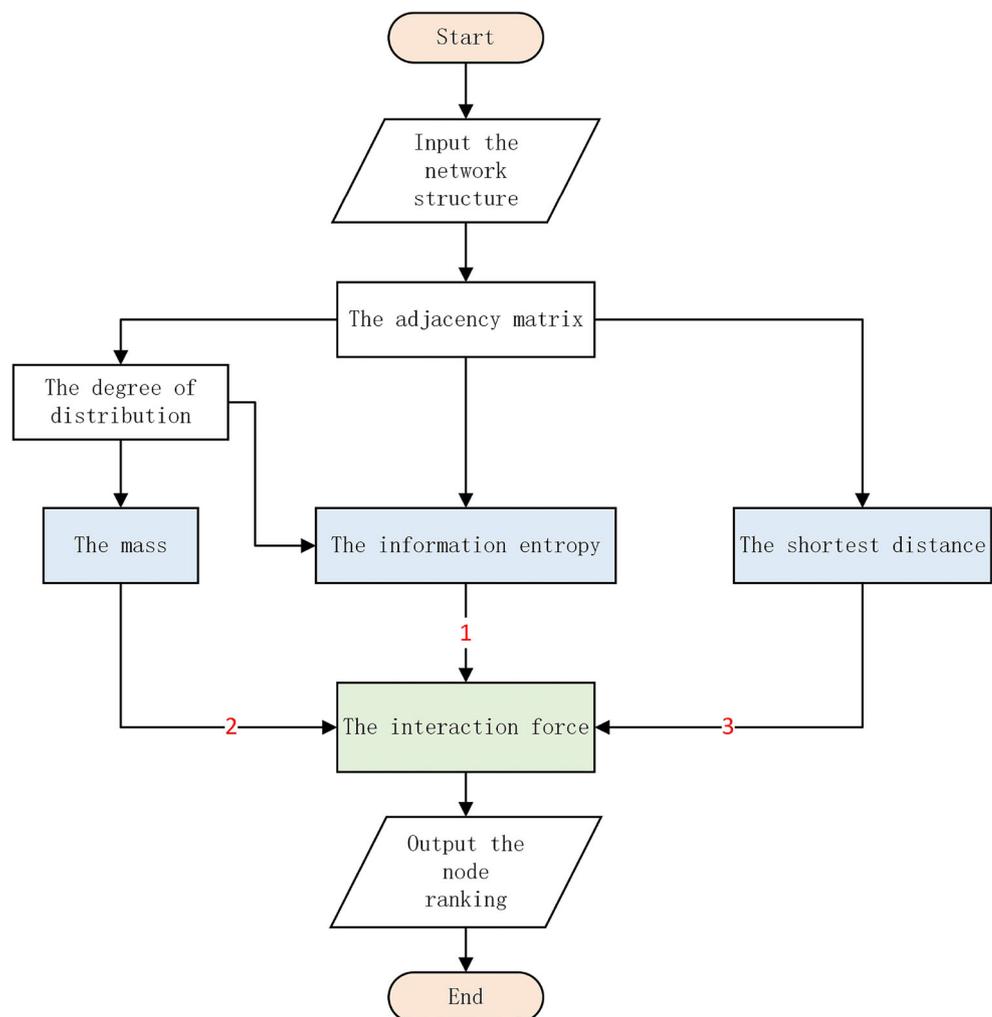
Input the mass, shortest distance, and weight of each node. Then the interaction force between the nodes is calculated. Output the ranking of the node influence in the complex networks.

We will introduce each step in detail in Section 3.2 to Section 3.6.

### 3.2 Step 1: construct the network

Given a real connected undirected network represented by  $G(V, E)$ , where  $V$  and  $E$  are the set of nodes and edges respectively. And the adjacency matrix  $A$  is used to store  $G$ , where the element  $a_{ij}$  in the matrix  $A$  represents the connection relationship between the node  $i$  and the node  $j$ .  $a_{ij} = 0$  means there is no connection between  $i$  and  $j$ . On the contrary,  $a_{ij} = 1$  means there is a connection between  $i$  and  $j$ .

**Fig. 1** The flow of the INEG algorithm. INEG has five steps, and the final result is composed of three factors



### 3.3 Step 2: calculate the degree distribution and the shortest distance

The adjacency matrix  $A$  of the network is given, then the degree of node  $i$  can be defined as follows:

$$k_i = \sum_j a_{ij} \tag{12}$$

where  $k_i$  denotes the degree of node  $i$ , and  $a_{ij}$  represents the connection between  $i$  and  $j$ . At the same time, the shortest distance between disparate pairs of nodes is defined as the number of edges on the shortest path connecting these two nodes. The shortest path is an ordered list ( $PAT H_{ij} = \{(i_0, i_1), (i_1, i_2), \dots, (i_{n-1}, i_n)\}$ ) containing the fewest connections, the number of which is  $L$ . The shortest distance  $d_{ij}$  is the length  $L$  of the shortest path. In this paper, Floyd’s algorithm [49] is used to calculate the shortest path in the network.

### 3.4 Step 3: Calculate the information entropy and get the weight

Each node in the network is affected by the global network and the local network. The local network of any node  $i$  in a complex network is denoted by  $L(N, D)$ , where  $N$  represents the set of nodes in the local network,  $D$  denotes the degree set of each node in  $N$ . As shown in Fig. 2, the local network of node 1 is composed of blue nodes.  $N = \{1, 2, 3, 4, 5\}$ ,  $D = \{4, 2, 3, 3, 4\}$ . We use  $P(i)$  to represent the probability set of  $i$ , which is defined as follows:

$$P(i) = [p(i, 1), p(i, 2), \dots, p(i, s), \dots, p(i, m)] \tag{13}$$

The local area network of each node includes its neighbor nodes and itself, so the scale  $m$  of the probability set is the

number of node sets  $N$ . The local probability set  $p(i, s)$  of node  $i$  is defined as follows:

$$p(i, s) = \frac{k_s}{\sum_{u=1}^m k_u} \tag{14}$$

For node 1,  $m = 5$ , and the probability set of node 1 is  $P(1) = [\frac{4}{16}, \frac{2}{16}, \frac{3}{16}, \frac{3}{16}, \frac{4}{16}]$ .

In the traditional gravity model, only the influence caused by the degree and distance of the node is considered, but the neighborhood information of the node is not considered. A node’s neighbor degree distribution is used to represent its neighborhood information. If a node’s neighbor degree distribution is more uniform, it will have less influence on the deviation of information propagation. For example, there are two nodes in the network with the same degree. If the neighbor degree distribution of a node is more uniform, the uncertainty in the direction of information diffusion will be greater. This means that it can spread the information better, and obviously its influence is greater. For another node whose neighbor degree distribution is not uniform, it has a higher probability of propagating information to nodes with a smaller degree. Therefore, it is less important. To quantify the degree distribution in the local network of nodes, the information entropy is introduced. It is well known that information entropy is used to measure the chaotic degree of distribution. The more uniform the distribution is, the greater the information entropy is. Information entropy, the expectation concerning the total amount of information, is defined as follows:

$$E_i = -\frac{1}{\log(n)} \sum_{p(i,s) \in P(i)} p(i, s) \log(p(i, s)) \tag{15}$$

The function of the coefficient  $\frac{1}{\log(n)}$  is to make the data distributed between 0 and 1. In Fig. 2, the entropy  $E_1$  representing node 1 is the relative score of node 1 in the local network.  $E_1 = -\frac{1}{\log(5)} \times (\frac{4}{16} \log \frac{4}{16} + \frac{2}{16} \log \frac{2}{16} + \frac{3}{16} \log \frac{3}{16} + \frac{3}{16} \log \frac{3}{16} + \frac{4}{16} \log \frac{4}{16})$ .

### 3.5 Step 4: calculate the mass of the nodes

Traditionally, a node’s K-shell value equal to  $x$  indicates that it forms a complete subgraph with  $x$  other nodes in the network (this complete subgraph contains  $x + 1$  nodes). A larger  $x$  denotes that the complete subgraph formed by it is larger, which means that it has a higher connection density. That is to say, a node with a larger K-shell value is more important. However, the hierarchical results of the traditional K-shell algorithm are rough, and nodes in the same shell cannot be distinguished. Therefore, a mixed location index  $km$ , which is the sum of the  $b$  index and the  $f$  index ( $km = b + f$ ), is proposed. Figure 3 visually shows the  $b$  and  $f$  values of the nodes in the network. The specific calculation process is as follows.

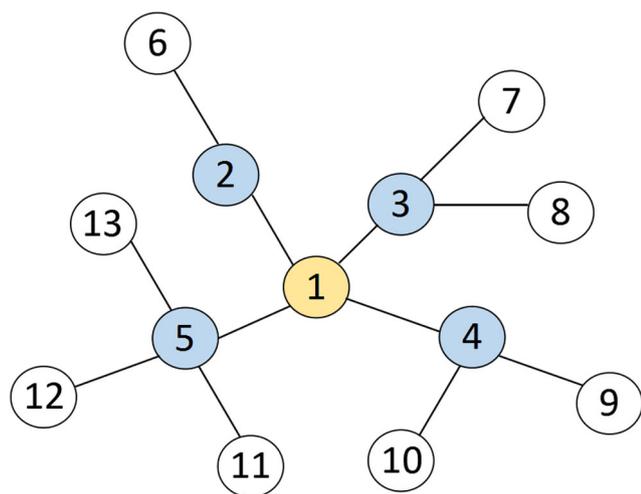
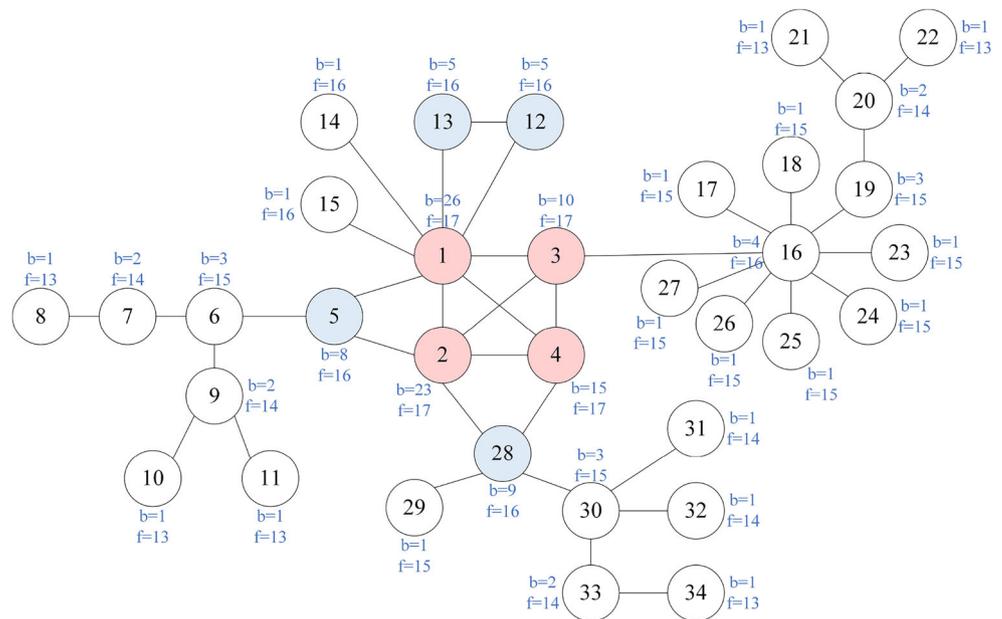


Fig. 2 An instance. The local network about node 1

**Fig. 3** The  $b$  and  $f$  of nodes in a network. The red nodes indicate that they are in shell 3, the blue nodes indicate that they are in shell 2, and the white nodes indicate that they are in shell 1. The red part is called the network core



1. The calculation process of the  $b$  index

The  $b$ -index is used to measure the distance of a node from the periphery of the system. A global iteration counter is introduced based on the traditional K-shell. In order to highlight the importance of the core part compared to the periphery and to distinguish the nodes as much as possible, inheritance rules are proposed. Whenever the algorithm runs to a new shell (ie  $shell + 1$ ), the  $b$  value of the node in the new shell will first inherit the  $b$  value from the neighbor nodes in the previous shell. Then, the number of global iterations to delete itself is counted. That is to say, a node's  $b$  value is equal to the  $b$  of its neighbors in the previous shell plus its own global iteration  $b$ . The global iteration  $b$  index actually refers to the number of iterations to delete node  $i$ . The allocation steps for the  $b$ -index are described below. Initialize  $shell = 1$  and  $b = 1$ . In the first step, the node with degree equal to  $shell$  ( $shell = 1$ ) is removed from the graph and assigned  $b = 1$ . When there are nodes in the graph with degree equal to  $shell$ , repeat this step, but increase  $b$  by 1. The second step, first increment the  $shell$  by 1. Next, the initial  $b$ -values of all nodes in the new shell are equal to the sum of their neighbors'  $b$ -values in the previous shell. Nodes with degree equal to  $shell$  are then again removed from the graph, and the global  $b$  value for a given counter is added to the  $b$  of the node. Each time a round of node deletion is performed, the  $b$  index is incremented by 1. When there are nodes in the graph with degree less than or equal to  $shell$ , this process is repeated until all nodes are removed. In this approach, nodes with larger  $b$  are closer to the core of the graph. A node whose  $b$ -value is

not less than any of its neighbors'  $b$ -values is considered a system core, and a system can have multiple cores.

2. The calculation process of the  $f$  index

The  $f$ -index measures how close a node is to the system core, and is primarily used to try to improve those nodes that are close to the system core but have lower  $b$ . The  $f$ -index depends on the  $b$ -index assigned in the previous step, and the specific process is as follows. In the first step, the nodes located in the core of the graph are identified. Then, the geometric mean of the  $b$ -index of these nodes is obtained, which is the  $f$ -index value of these nodes. In the second step, the traversal starts from the core of the graph. All nodes with the highest  $f$  value in the graph assign values  $(f - 1)$  to the  $f$ -index of their undeleted neighbor nodes. After the assignment, remove from the graph those nodes with the largest  $f$ . In the third step, this process is repeated until all nodes in the graph are deleted.

Finally, the mass of node  $i$  introducing location information is expressed as follows:

$$M_i = e^{\frac{km_i - km(\min)}{km(\max) - km(\min)}} \times k_i \tag{16}$$

where  $km_i$  represents the mixed position index of node  $i$ .  $km(\max)$  and  $km(\min)$  represent the maximum and minimum values of the mixed location indicators in the graph, respectively.  $\frac{km_i - km(\min)}{km(\max) - km(\min)}$  stands for normalization of mixed location metrics. Obviously, when  $km_i = km(\min)$ , the location information will have no effect on the mass of the node. This means that when the mixed location metric is minimal, the main factor affecting node mass is degree.

### 3.6 Step 5: calculate the interaction force

In the mechanical model, the gravity is proportional to the product of their masses and inversely proportional to the square of their distance. By analogy with the mechanical model and combining the above steps, the interaction force at node  $i$  is defined as:

$$INEG_i = \sum_{d_{ij} \leq 0.5 \langle d \rangle} E_i \times \frac{M_i \times M_j}{d_{ij}^2} \quad (17)$$

where  $d_{ij}$  indicates the shortest distance between  $i$  and  $j$ .  $R = 0.5 \langle d \rangle$ , which follows the setting principle of (7). In general, the generalized mechanics model based on information entropy consists of three parts: the mass, the shortest distance, and the information entropy weight of the node itself.

## 4 Experiments

In this section, we conduct three different experiments on eleven real networks and compare them with ten different methods in order to test the effectiveness of INEG.

### 4.1 Description of the data set

Eleven networks in the real world are selected, including collaboration networks (Jazz [50], NS [51], and GrQc [52]), communication networks (Email [53] and EEC [54]), social networks (PB [55], Facebook [56], and WV [57]), transportation networks (USAir [58]), infrastructure networks (Power [59]) and technical network (Router [60]).

Jazz [50] describes the different collaborative relationships between jazz musicians. NS [51] shows a collaborative network of scientists. In Jazz and NS, a node denotes a musician and a scientist respectively, and an edge represents the cooperation between any two musicians or scientists. GrQc [52] covers collaborative scientific papers on general relativity and the quantum universe from the arXiv preprint site. Email network [53] reveals the email exchange among members. EEC [54] depicts a network of European research members exchanging mail. For Email and EEC, a node indicates a user. If two users communicate with each other, there will be an edge between a pair of nodes. PB [55] represents the blogs' connection relationship in the USA. The Facebook social circle is described in Facebook [56]. The dataset includes node features (profiles), circles, and ego-networks. WV [57] summarizes the network of Wikipedia voting issues. In Wikipedia, in order for a user to become an administrator, an administrative request needs to be issued. Then, the Wikipedia community decides who is promoted to be an administrator through public voting. USAir [58] is the American transportation aviation network, which describes the major airports and corresponding airlines in the United States. A node indicates an airport, and an edge denotes airlines between different airports. Power [59] represents the power grid in the western United States. For Power, the node represents a generator, converter, or substation. And an edge indicates a power line. Router [60] denotes a symmetrical snapshot of the Internet structure at the autonomous system level. These data sets can be downloaded from <http://snap.stanford.edu/data/> and <http://vlado.fmf.uni-lj.si/pub/networks/data/>.

The specific information of these networks is shown in Table 1, where  $N$  represents the node,  $E$  denotes the edge.

**Table 1** Some basic topological characteristics of eleven selected real-world networks

Type	Network	$ N $	$ E $	$\langle d \rangle$	$\langle k \rangle$	C
Small-scale	Jazz	198	2742	2.2350	27.6970	0.6175
	USAir	332	2126	2.7381	12.8072	0.6252
	NS	379	914	6.0419	4.8232	0.7412
	EEC	986	16064	2.5869	32.5842	0.4505
	Email	1133	10903	3.6060	9.6240	0.2202
	PB	1222	16714	2.7375	27.3552	0.3600
Large-scale	Facebook	4039	88234	3.6925	43.6910	0.6170
	GrQc	4158	13422	6.0494	6.4560	0.6648
	Power	4941	6594	18.9892	2.6691	0.1065
	Router	5022	6258	2.4922	6.4488	0.0329
	WV	7066	100736	3.2475	28.5129	0.2090

$\langle d \rangle$  shows the average value of the shortest distance in the network,  $\langle k \rangle$  represents the average value of the network.  $C$  is the clustering coefficient of the network.

## 4.2 Simulation

### 4.2.1 Top-ten nodes

The SI model is the simplest infectious disease model, which can be used to evaluate the spreading ability of each node and reveal the influence of nodes. In the SI model [61], the nodes of the network are segmented into two states: susceptible state (Class  $S$ ) and infectious state (Class  $I$ ). The susceptible nodes refer to the nodes that don't have the disease, but lack immunological competence, and are prone to be infected after contact with the infected nodes. Infected nodes refer to nodes infected with infectious diseases, which have a probability to affect their neighbor's susceptible nodes and turn them into infected state. Initially, one selected node (called the seed) is in the  $I$  state while the other nodes are in the  $S$  state. When the process begins, the initially infected node has an ability to infect its surrounding nodes in  $S$  state with the probability  $\beta$ . Castellano and Pastor-Satorras [62] suggest that the infectious rate  $\beta$  is slightly above the epidemic threshold probability  $\beta_c$  ( $\beta_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$ ). The standard SI model sequence for each network is obtained according to the epidemic threshold probability.

The top-ten nodes of the twelve measures on eleven different networks are demonstrated in Table 2. Because different methods evaluate importance from different aspects of the network structure, the ranking lists they obtained would be different. The ranking list of the SI model, which represents the standard sequence of the network, shows the top-ten key nodes in each network. For a clearer display, the same nodes in the ranking list for each measure and the standard sequence are shown in red, and different nodes are shown in black. For example, if node 1 is identified by the standard SI model and also identified by the DC, node 1 in the DC is marked in red. The effectiveness of each method is demonstrated by comparing each measure with the standard sequence. Compare each measure to the standard sequence to demonstrate their effectiveness. A method is more efficient if it has more nodes in common with the standard sequence.

It can be seen from Table 2 that the effectiveness of INEG is generally relatively high in all networks. In Jazz, INEG has six nodes that are the same as the standard sequence. It is second only to DC, CC, and GGC. For USAir, the ranking list for each method is almost identical to the standard sequence. The recognition performance of INEG is only inferior to GGC and LID in NS. Except for CC, BC, LID, and FLD, other measures have eight nodes that are the same as the standard sequence in EEC. For Email, CC and LID

are tied for first place in performance. In PB, DC and PC are identical to the standard sequence. For Facebook, EC has the worst performance. For GrQc, the capabilities of INEG, DC, GC, and WGC are tied for first place. INEG's performance ranks third in Router. For WV, LID identifies two vital nodes, which proves its performance is not good.

### 4.2.2 Individuation

The Individuation experiment, which is used to count the frequency of nodes under different rankings, is defined as follows:

$$Frequency = \frac{n_u}{n} \quad (18)$$

where  $n_u$  represents the number of nodes with unique scores, and  $n$  denotes the total number of nodes in the network. The frequency metric depicts the proportion of nodes with unique scores in the network. The experiment is used to evaluate the ability to distinguish nodes of each method. An approach with a larger rank and a smaller frequency means that it scores nodes more spread out. Therefore, this method can better distinguish the importance of nodes in the network. For example, if a method assigns the same score to all nodes in extreme cases, the ability of this method to distinguish nodes is extremely poor. This means that all nodes are ranked first and the method cannot identify crucial nodes. The specific values of the experiment are shown in Table 3. Red indicates the best performance, blue represents second, and green denotes third. What's more, the macro effects on ten networks are shown in Figs. 4 and 5.

For Jazz, the highest ranking and lowest frequency are obtained by INEG. INEG's capability is only inferior to WGC in USAir. INEG's performance ranks second in NS. For EEC and PB, INEG defeats other methods. WGC and INEG tied for first place in Email. For Facebook, GrQc, and Router, the performance of INEG is only inferior to WGC. In WV, the best performance is demonstrated by INEG.

### 4.2.3 Kendall's tau coefficient

The Kendall correlation coefficient [63] is named after Maurice Kendall, and often uses the Greek letter  $\tau$  to indicate its value. It is a statistical value used to measure the correlation between two random sequences. Two random sequences  $X$  and  $Y$  with  $N$  elements are given, and their  $i$ th ( $1 \leq i \leq N$ ) elements are represented by  $X_i$  and  $Y_i$  respectively. At the same time, the corresponding elements form a set  $XY$ , which contains elements  $(X_i, Y_i)$  ( $1 \leq i \leq N$ ). Considered any pair of elements  $(X_i, Y_i)$  and  $(X_j, Y_j)$  in the set, when  $X_i > X_j$  and  $Y_i > Y_j$  or  $X_i < X_j$  and  $Y_i < Y_j$  occurs, this pair of elements is regarded consistent. When

**Table 2** The similarity analysis of the top-ten nodes ranked by ten diverse approaches in ten distinguishing networks

Rank	Standard	Classic					Similar			State-of-the-art		Proposed
	SI model	DC	CC	BC	EC	PC	GC	WGC	GGC	LID	FLD	INEG
<b>Jazz</b>												
1	8	8	8	8	100	8	8	100	8	8	8	100
2	100	100	100	155	4	100	100	8	100	131	131	4
3	194	4	131	100	8	131	4	4	131	194	194	8
4	131	131	194	186	131	4	131	131	4	32	32	131
5	69	194	69	131	80	186	80	80	194	79	69	80
6	4	80	4	136	129	136	194	129	69	69	53	5
7	59	129	32	127	5	194	129	194	162	150	111	129
8	162	69	53	60	194	69	69	5	53	155	186	32
9	111	162	111	28	69	28	5	69	129	173	79	194
10	67	77	162	69	53	175	53	53	80	4	49	69
<b>USAir</b>												
1	118	118	118	118	118	118	118	118	118	118	118	118
2	261	261	261	8	261	261	261	261	261	67	261	261
3	255	255	67	261	255	182	255	255	255	261	67	255
4	182	152	255	201	182	152	182	182	182	201	255	182
5	166	182	201	47	152	255	152	152	152	47	201	166
6	67	230	182	182	230	230	230	230	166	255	182	152
7	152	166	47	255	112	166	166	166	230	166	166	67
8	201	67	166	152	67	201	67	67	67	248	47	112
9	230	112	248	313	166	67	112	112	201	182	248	230
10	248	201	112	13	147	8	147	147	112	112	112	176
<b>NS</b>												
1	26	4	26	26	4	26	4	4	4	51	51	5
2	5	5	95	51	5	4	5	5	26	26	95	4
3	4	26	51	169	16	5	26	16	5	52	231	16
4	51	16	231	95	15	95	16	15	16	67	26	26
5	231	67	100	67	45	67	15	45	51	95	52	231
6	1	70	52	5	46	16	51	1	95	5	5	15
7	52	95	5	231	47	32	95	46	231	169	169	95
8	95	15	44	100	176	51	231	47	67	16	100	70
9	16	32	234	44	177	8	67	176	52	23	170	51
10	169	51	297	66	250	70	70	177	169	231	76	67
<b>EEC</b>												
1	161	161	161	161	161	161	161	161	161	161	161	161
2	122	122	83	87	122	122	122	122	122	83	83	122
3	83	83	122	6	83	83	83	83	83	122	122	83
4	63	108	108	83	108	108	108	108	108	108	108	108
5	87	87	63	122	63	87	63	63	63	63	63	63
6	108	63	87	108	435	63	87	435	87	87	87	250
7	435	435	435	14	250	6	435	250	435	435	435	184
8	184	14	167	378	184	14	250	87	167	167	167	167
9	6	167	250	63	87	167	167	184	184	250	250	435
10	130	184	65	65	167	435	184	167	250	65	65	87

Table 2 (continued)

Rank	Standard	Classic					Similar			State-of-the-art		Proposed
	SI model	DC	CC	BC	EC	PC	GC	WGC	GGC	LID	FLD	INEG
Email												
1	333	105	333	333	105	105	105	105	105	105	333	105
2	105	333	23	105	16	23	42	16	333	333	23	42
3	23	16	105	23	196	333	16	196	42	23	42	333
4	42	23	42	578	204	41	333	42	23	41	105	41
5	52	42	41	76	42	42	23	333	41	42	76	299
6	3	41	76	233	49	233	196	23	16	233	468	23
7	233	196	233	135	56	16	41	204	196	76	41	16
8	332	233	52	41	116	355	76	49	233	52	233	434
9	378	21	135	355	333	21	3	3	76	135	52	196
10	135	76	378	42	3	24	233	116	52	3	378	233
PB												
1	127	127	838	672	127	672	127	127	838	127	838	127
2	838	838	127	127	48	127	838	48	127	838	890	48
3	497	672	497	768	497	768	48	497	497	497	673	497
4	672	48	48	838	566	838	497	838	48	48	592	566
5	566	497	890	497	283	497	566	566	672	672	639	283
6	768	768	566	1178	147	48	672	283	566	768	692	838
7	48	1006	768	48	838	1006	1006	147	1006	566	497	147
8	922	566	922	782	84	922	922	498	922	890	566	384
9	1178	922	1178	922	384	1178	890	253	768	922	101	119
10	1006	1178	672	566	498	566	830	384	890	1006	127	498
Facebook												
1	108	108	108	108	1913	3438	108	1913	108	108	108	1913
2	1685	1685	59	1685	2267	108	1913	2348	1913	1685	429	108
3	1913	1913	429	3438	2207	1685	1685	2544	1685	1913	564	2348
4	3438	3438	564	1913	2234	1	2348	2267	3438	484	1466	2544
5	1942	1	1685	1086	2465	1913	2544	2234	2544	349	1719	2267
6	2348	2544	172	1	2143	349	2267	1986	2348	415	1578	1986
7	2544	2348	349	699	2219	687	1986	2143	1889	1	607	2234
8	1719	1889	484	568	2079	3981	2234	2207	1801	429	367	2207
9	484	1801	415	59	2124	415	1889	2411	1664	377	518	2143
10	1585	1664	377	429	1994	699	2143	2219	1353	476	527	2219
GrQc												
1	3348	3348	2129	2129	3348	2174	3348	3348	3348	3261	2129	1924
2	1924	3388	2205	1525	437	2129	1924	3388	3614	2139	3956	3388
3	1065	1924	1525	2221	1924	2139	3388	1924	1924	3956	3261	1065
4	2754	3614	2754	1232	3419	1525	3614	3614	3388	3534	1525	3348
5	3614	1065	429	2139	1554	433	1065	1554	2129	2036	2036	3614
6	3388	1554	3348	817	2277	3388	2754	1065	2754	1260	429	3075
7	3715	3419	1947	2205	4012	1232	1554	3419	1065	1678	1947	437
8	437	2754	3956	433	1267	1012	437	2754	2139	461	1678	1554
9	3075	437	1924	2174	2251	3614	3419	437	1554	433	2754	2754
10	1554	3075	3614	2754	1995	3348	3075	3075	2978	2892	2139	251

**Table 2** (continued)

Rank	Standard	Classic					Similar			State-of-the-art		Proposed
	SI model	DC	CC	BC	EC	PC	GC	WGC	GGC	LID	FLD	INEG
<b>Router</b>												
1	3668	3670	3668	3668	3369	3670	3369	3369	3668	3668	3668	3326
2	3373	3639	3373	3373	3373	3639	3326	3338	3670	4	3373	3373
3	3326	3338	3667	1480	3339	3338	3668	3339	3326	228	3667	3369
4	3729	1453	3729	3326	3326	1453	3670	3326	3369	4875	3729	3339
5	3727	3369	3666	3729	3342	2624	3373	3373	3338	4912	3666	3668
6	3369	3339	623	3667	3325	3675	3339	3342	3339	243	3727	3338
7	3670	3326	3326	4059	3352	3369	3338	3334	3639	1123	1480	3670
8	3325	2624	3325	3597	3338	3672	3639	3354	3373	4683	3326	3729
9	3339	3373	3727	1449	3334	3339	3729	3331	3729	3853	3325	3342
10	3338	3675	3324	623	3327	3644	3727	3325	3727	3958	3687	3639
<b>WV</b>												
1	2355	2355	2355	2355	2355	2355	2355	2355	2355	2355	2355	2355
2	703	703	703	9	703	9	1428	1428	703	4890	703	1428
3	432	9	432	432	1428	703	703	703	1428	3644	432	1067
4	1428	1428	1428	3644	1067	432	1067	1067	432	677	1067	703
5	1067	432	1067	1428	2470	3644	432	432	1067	13	1428	432
6	9	1067	1262	703	432	1428	9	2470	9	2203	66	9
7	2470	2470	9	1067	3059	1067	2470	9	2470	4522	677	1055
8	1262	1262	1055	13	9	2470	1262	3059	1262	2719	4890	2470
9	1055	1055	2470	1262	1055	13	1055	1262	1055	3872	13	1262
10	677	4890	2282	2057	1262	1262	3059	1055	3059	294	294	4890

$X_i > X_j$  and  $Y_i < Y_j$  or  $X_i < X_j$  and  $Y_i > Y_j$  occurs, this pair of elements is deemed inconsistent. When  $X_i = X_j$  and  $Y_i = Y_j$ , this pair of elements is considered neither consistent nor inconsistent. Therefore, Kendall correlation coefficient is defined as follows:

$$\tau = \frac{n_c - n_d}{0.5n(n - 1)} \tag{19}$$

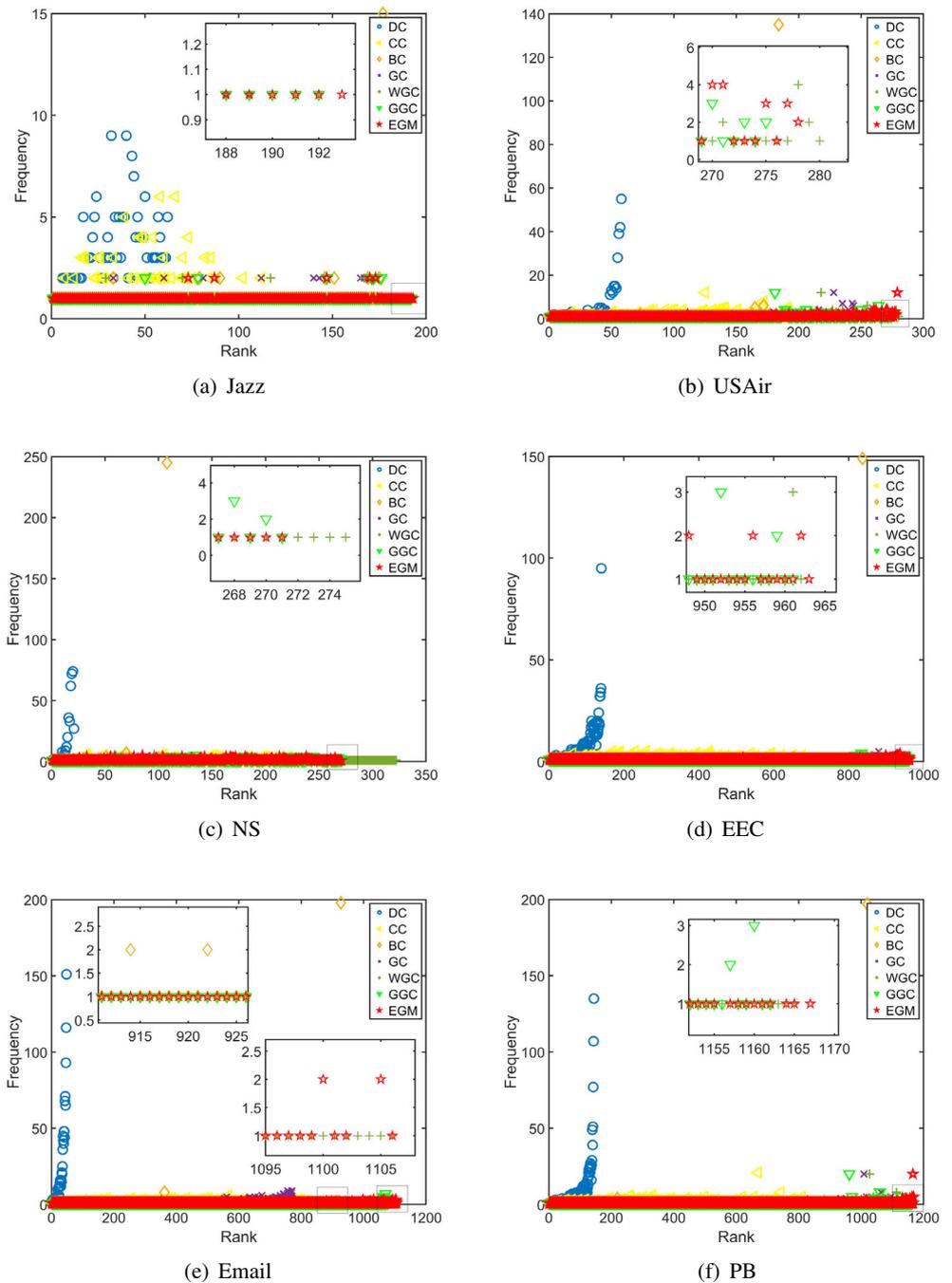
where  $n_c$  indicates the number of identical pairs. On the contrary,  $n_d$  represents the number of discordant pairs. And  $n$  denotes the length of the sequence. Its value lies between  $-1$  and  $1$ .  $\tau = 1$  indicates the two sequences have totally the same rank, while  $\tau = -1$  is the opposite.

The higher the  $\tau$  is, the more precise the centrality measure is. We use Kendall’s tau coefficient to measure the correlation between the ranked list of different methods and

**Table 3** Individualized ranking of seven measures

Type	Network	DC	BC	CC	GC	WGC	GGC	INEG
Small-scale	Jazz	0.31313	0.89394	0.64141	0.95455	0.96970	0.96970	0.97475
	USAir	0.17470	0.55422	0.58133	0.78614	0.84337	0.82831	0.84036
	NS	0.05541	0.28496	0.60158	0.69921	0.84960	0.71504	0.71504
	EEC	0.14199	0.84888	0.68560	0.94828	0.97566	0.97465	0.97667
	Email	0.04237	0.81818	0.74051	0.68049	0.97617	0.95322	0.97617
	PB	0.11784	0.83470	0.67430	0.90671	0.95336	0.95090	0.95581
Large-scale	Facebook	0.05620	0.86927	0.30057	0.91681	0.97673	0.95791	0.96014
	GrQc	0.01563	0.36989	0.66955	0.78523	0.88793	0.79966	0.84055
	Router	0.01055	0.12405	0.29669	0.35743	0.42433	0.37017	0.41876
	WV	0.04246	0.64704	0.62780	0.62270	0.81885	0.81475	0.82154

**Fig. 4** For six small-scale networks, the frequency of nodes obtained by different approaches. The X-axis denotes the ranking of the nodes, and the Y-axis represents the number of nodes

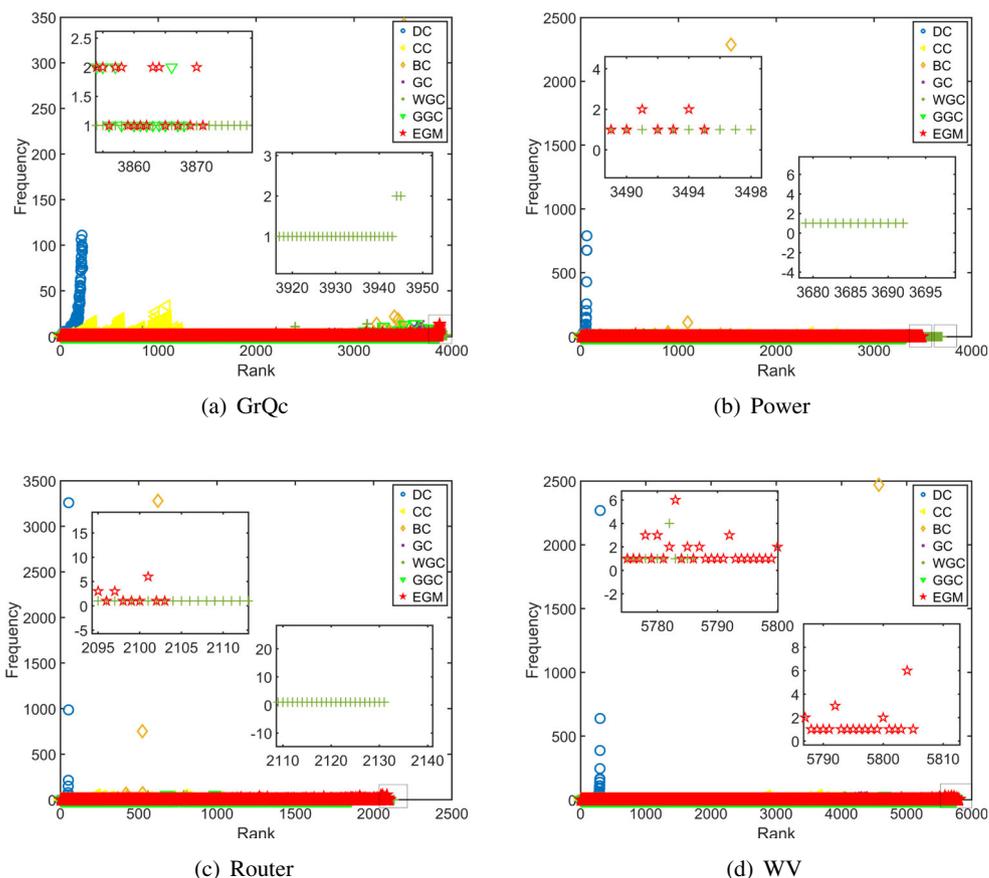


the ranked list of standard SI model. Time  $t$  is set to 10. The quantity of infected nodes in ten steps ( $F(10)$ ) is used to represent the infection ability of each node.

The results in the small-scale networks are indicated in Fig. 6. INEG defeats five methods in Jazz. The trend of INEG is roughly the same as that of WGC in USAir. In addition, INEG is superior to similar methods based on gravity models like GGC. For NS, the trends of GC and

INEG are almost the same. What's more, INEG ranks first when  $0.1 \leq \beta \leq 0.12$ . INEG's performance is similar to WGC in EEC. In addition, the highest  $\tau$  value is obtained by INEG when  $0.1 \leq \beta \leq 0.13$ . For Email, the performance of INEG is the best when  $\beta = 0.19$ . For PB, INEG and GC have almost the same trend. In particular, they almost defeat other approaches. At the same time, Fig. 7 demonstrates the results in the large-scale networks. The  $\tau$  value obtained

**Fig. 5** For four large-scale networks, the frequency of nodes obtained by different approaches. The X-axis denotes the ranking of the nodes, and the Y-axis represents the number of nodes



by INEG is almost the same as that obtained by GC in GrQc. Especially when  $0.1 \leq \beta \leq 0.11$ , INEG ranks first. For Power, the trend shown by INEG is almost the same as that shown by GC. And their  $\tau$  values are higher than those acquired by other methods. In Router, INEG ranks first when  $0.12 \leq \beta \leq 0.15$ . For WV, INEG acquires the highest  $\tau$  value when  $0.16 \leq \beta \leq 0.21$ .

## 5 Applications

### 5.1 Explore the application of immune strategies

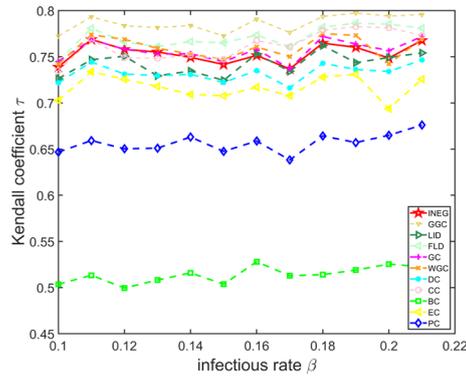
Immunization strategy is one of the most basic structural problems in network science. Its main goal is to effectively control the spread of diseases by vaccinating some people. The method proposed in this paper can mine this part of the key population in a large-scale network, thereby reducing the consequence caused by the virus and achieving a better immune effect.

The SI model experiment is used to measure the immune effect of each method. The top-ten nodes of each method

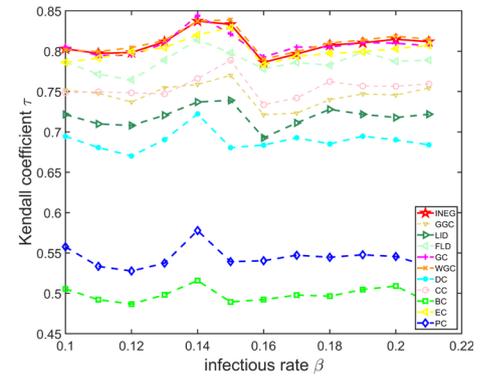
are selected as their initial infection nodes, and  $\beta$  is set to the epidemic threshold probability. The ranking of the top ten nodes has been given in the Top-ten nodes experiment. We conducted one hundred experiments and the results are revealed in Figs. 8 and 9. In a certain period of time, a method that gets more infection nodes indicates that the nodes it initially chooses are more influential. In other words, the performance of this approach is better.

The experimental results in the small-scale networks are revealed in Fig. 8. For Jazz, the most infected nodes are demonstrated by INEG when  $t \geq 31$ . Meanwhile, INEG exhibits the best performance when all approaches reach stability. In USAir, the strongest infectious ability and the perfect stability are proved by INEG when  $t \geq 26$ . Faster infection rate is revealed by INEG in the initial stage of the NS. All methods show almost the same trend in EEC. In addition, INEG ranks fifth. For Email, CC has the strongest infection ability which surpasses other methods. For PB, the most infected nodes are acquired by INEG when  $t \geq 45$ . Figure 9 shows the experimental results in the large-scale networks. The largest number of nodes is obtained by LID in GrQc. In Power, INEG outperforms

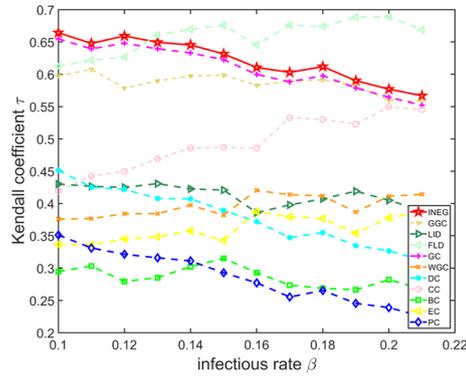
**Fig. 6** Kendall's tau coefficient  $\tau$  between the infection ability acquired by the SI model and the other methods for six small-scale networks. The X-axis represents the infection rate, the Y-axis indicates the Kendall's tau coefficient, and each point represents the value of  $\tau$  under the corresponding infection rate



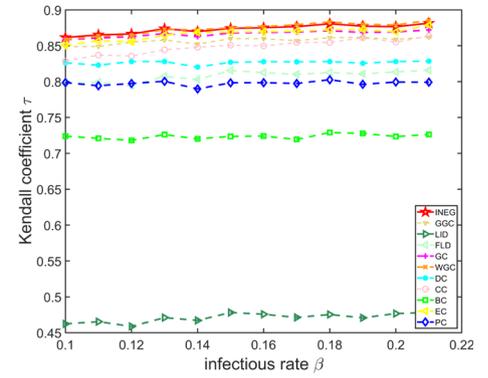
(a) Jazz



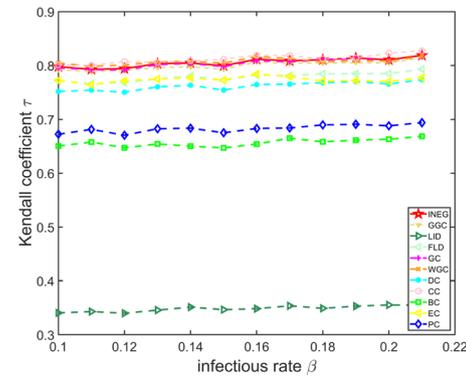
(b) USAir



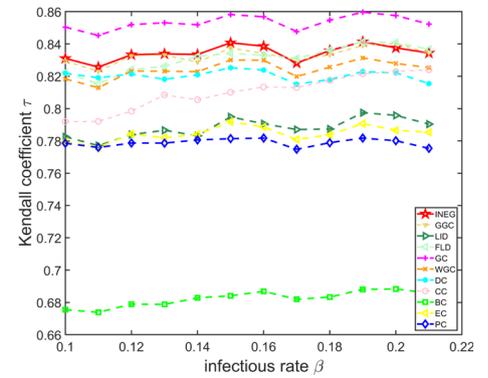
(c) NS



(d) EEC

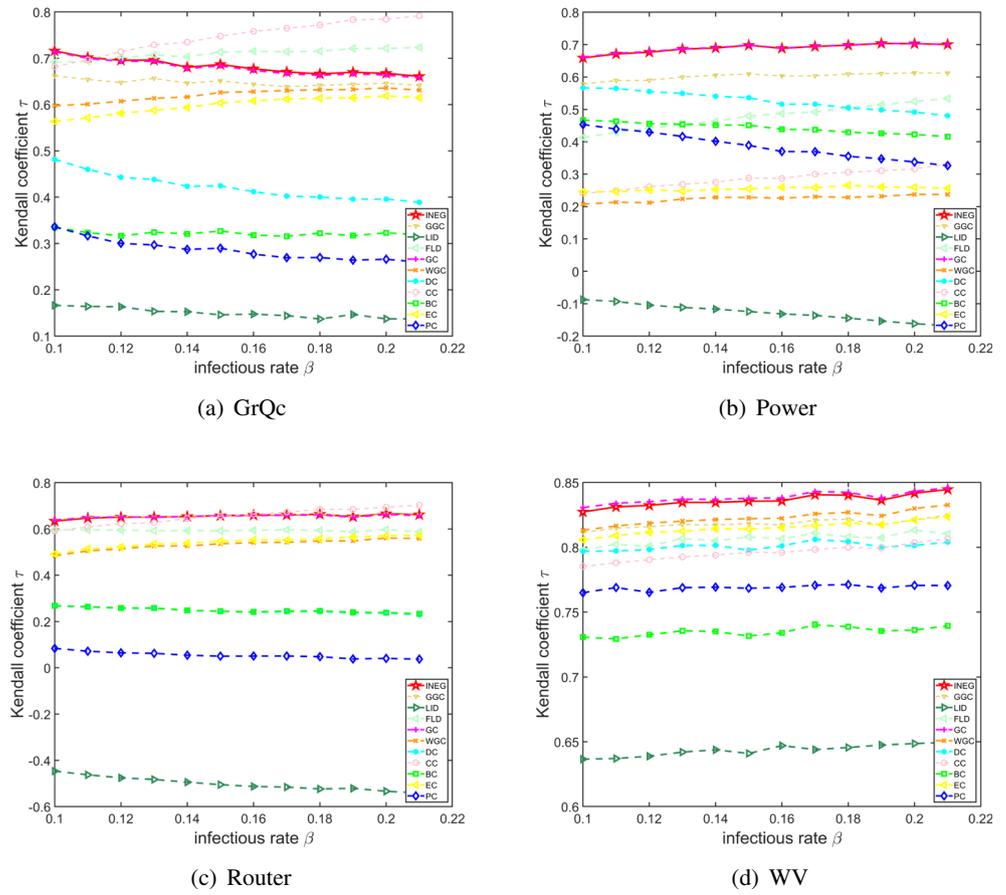


(e) Email



(f) PB

**Fig. 7** Kendall's tau coefficient  $\tau$  between the infection ability acquired by the SI model and the other methods for four large-scale networks. The X-axis represents the infection rate, the Y-axis indicates the Kendall's tau coefficient, and each point represents the value of  $\tau$  under the corresponding infection rate



other methods almost the entire time period. For Router, INEG's performance ranks fourth when  $t = 46$ . For WV, INEG defeats other measures almost the entire time period.

**5.2 Other applications**

**5.2.1 Applications for finding super-spreaders**

With the outbreak of COVID-19, the World Health Organization has increased its emphasis on epidemic control. During the entire process of the outbreak, the discovery and isolation of 'super-spreaders' has become a key measure to control the epidemic. Super-spreader, which is defined as the transmission of the virus to more than ten patients, is the key node in the epidemic transmission network. The method proposed in this paper is used to find super-spreaders in epidemic transmission models.

**5.2.2 Figure pooling application**

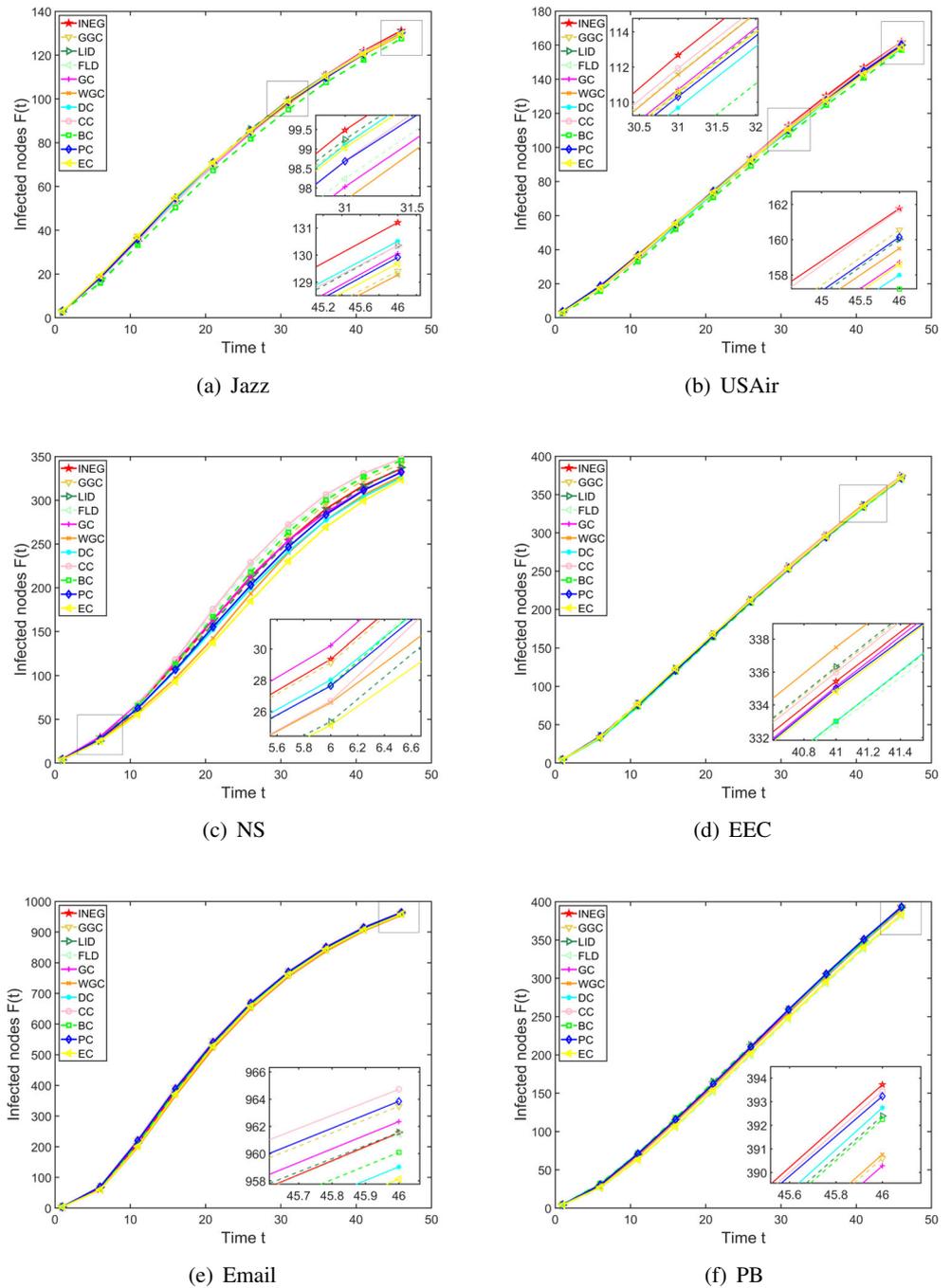
Pooling in graph classification tasks is a fundamental problem to be solved by graph neural networks. Graph

neural network is a deep learning method based on graph structure, which is used for processing data structure in non-Euclidean space. Graph classification, which is a typical task in graph neural networks, is used to learn a graph classification model from the figure to the corresponding label. Graph classification focuses on the global information of graph data, including structure information and attribute information of each node. In convolutional neural networks, hierarchical pooling is an important method used to gradually extract global information. As one of the main pooling mechanism methods, Top-K is a process of continuously discarding nodes for graphs with irregular structures. Important nodes that account for a fixed proportion are looked for to replace global information, and then the remaining nodes are ignored. The approach proposed in this paper can be applied to find crucial nodes with a fixed ratio.

**6 Conclusion**

The paper proposes a generalized mechanics model based on information entropy to evaluate the influence of nodes in the network. The method not only considers

**Fig. 8** The number of infected nodes for dissimilar initial nodes (top-10 nodes) acquired by different approaches in six small-scale networks. The X-axis indicates the experimental simulation time  $t$ , the Y-axis denotes the number of infected nodes in the network at a specific time  $t$ , and each point reveals the number of infected nodes generated in the corresponding time

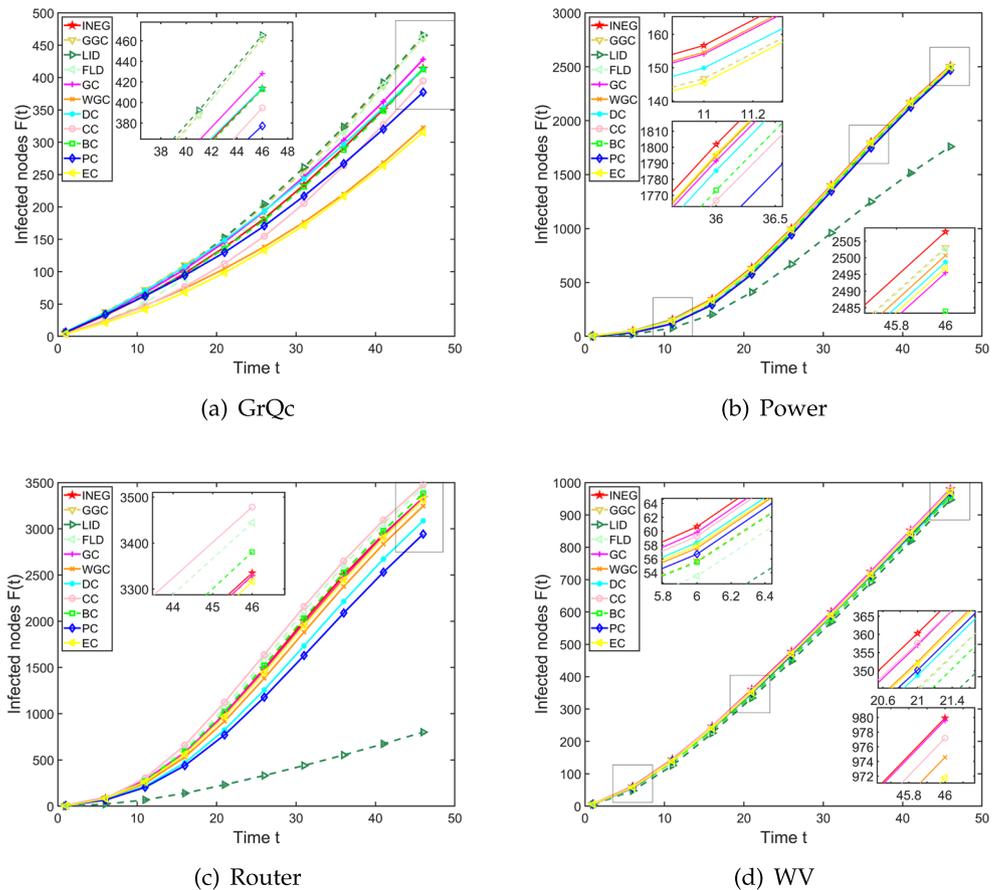


the global information by calculating the shortest path, but also considers the degree of the node itself in the local information. In addition, the influence received from neighbor nodes is quantified in detail, and the information entropy of the node is introduced to objectively assign its own weight to each node. At the same time, location information is introduced to form the mass. Local information and global information are better combined. The proposed method in this paper performs well in

different experiments on several real networks, which offers a broad prospect for identifying critical nodes.

Some ongoing works are as follows. Our approach only considers the influence of first-order neighbor nodes, while the influence of second-order or higher-order neighbor nodes are supposed to be expected. At the same time, the method in this paper is suitable for undirected networks. In the future, we will take into account improvements to the weighted network.

**Fig. 9** The number of infected nodes for dissimilar initial nodes (top-10 nodes) acquired by different approaches in four large-scale networks. The X-axis indicates the experimental simulation time  $t$ , the Y-axis denotes the number of infected nodes in the network at a specific time  $t$ , and each point reveals the number of infected nodes generated in the corresponding time



**Acknowledgements** The authors greatly appreciate the reviewers' suggestions and the editor's encouragement. This research is supported by the National Natural Science Foundation of China (No. 62003280), Chongqing Talents: Exceptional Young Talents Project (No. cstc2022ycjh-bgzxm0070), Natural Science Foundation of Chongqing, China (No. CSTB2022NSCQ-MSX0531), and Chongqing Overseas Scholars Innovation Program (No. cx2022024).

**References**

1. Newman ME (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256. <https://doi.org/10.1137/S003614450342480>
2. Huang C, Li M, Cao F, Fujita H, Li Z, Wu X (2022) Are graph convolutional networks with random weights feasible? *IEEE Trans Pattern Anal Mach Intell* :1–18
3. Hayashi T, Fujita H, Hernandez-Matamoros A (2021) Less complexity one-class classification approach using construction error of convolutional image transformation network. *Inf Sci* 560:217–234. <https://doi.org/10.1016/j.ins.2021.01.069>
4. Chen L, Deng Y, Cheong KH (2021) Probability transformation of mass function: A weighted network method based on the ordered visibility graph. *Eng Appl Artif Intell* 105:104438. <https://doi.org/10.1016/j.engappai.2021.104438>
5. Zareie A, Sakellariou R (2021) Minimizing the spread of misinformation in online social networks: A survey. *J Netw Comput Appl* 186:103094–0. <https://doi.org/10.1016/j.jnca.2021.103094>
6. Wen T, Gao Q, Chen Y-w, Cheong KH (2022) Exploring the vulnerability of transportation networks by entropy: A case study

- of Asia–Europe maritime transportation network. *Reliab Eng Syst Saf* :108578
7. Cimr D, Fujita H, Tomaskova H, Cimler R, Selamat A (2022) Automatic seizure detection by convolutional neural networks with computational complexity analysis. *Comput Methods Prog Biomed* 229:107277. <https://doi.org/10.1016/j.cmpb.2022.107277>
8. Tanveer M, Rashid AH, Ganaie M, Reza M, Razzak I, Hua K-L (2021) Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning. *IEEE J Biomed Health Inf* 26(4):1453–1463. <https://doi.org/10.1109/JBHI.2021.3083274>
9. Wang S-H, Govindaraj VV, Górriz JM, Zhang X, Zhang Y-D (2021) Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Inf Fusion* 67:208–229. <https://doi.org/10.1016/j.inffus.2020.10.004>
10. Zhang H, Zhong S, Deng Y, Cheong KH (2021) LFIC: Identifying influential nodes in complex networks by local fuzzy information centrality. *IEEE Trans Fuzzy Syst*:3284–3296. <https://doi.org/10.1109/TFUZZ.2021.3112226>
11. Lei M, Cheong KH (2022) Node influence ranking in complex networks: A local structure entropy approach. *Chaos Solitons Fractals* 160:112136. <https://doi.org/10.1016/j.chaos.2022.112136>
12. Zareie A, Sheikahmadi A, Jalili M (2020) Identification of influential users in social network using gray wolf optimization algorithm. *Expert Syst Appl* 142:112971. <https://doi.org/10.1016/j.eswa.2019.112971>
13. Wang X, Zhou W, Li R, Cao J, Lin X (2018) Improving robustness of interdependent networks by a new coupling strategy. *Phys A Stat Mech Appl* 492:1075–1080. <https://doi.org/10.1016/j.physa.2017.11.037>

14. Lei X, Yang X, Fujita H (2019) Random walk based method to identify essential proteins by integrating network topology and biological characteristics. *Knowl-Based Syst* 167:53–67. <https://doi.org/10.1016/j.knsys.2019.01.012>
15. Sheikhhahmadi A, Zareie A (2020) Identifying influential spreaders using multi-objective artificial bee colony optimization. *Appl Soft Comput* 94:106436. <https://doi.org/10.1016/j.asoc.2020.106436>
16. He Q, Sun L, Wang X, Wang Z, Huang M, Yi B, Wang Y, Ma L (2021) Positive opinion maximization in signed social networks. *Inf Sci* 558:34–49. <https://doi.org/10.1016/j.ins.2020.12.091>
17. Lin L-F, Li Y-M (2021) An efficient approach to identify social disseminators for timely information diffusion. *Inf Sci* 544:78–96. <https://doi.org/10.1016/j.ins.2020.07.040>
18. Zareie A, Sheikhhahmadi A, Jalili M, Fasaei MSK (2020) Finding influential nodes in social networks based on neighborhood correlation coefficient. *Knowl-Based Syst* 194:105580. <https://doi.org/10.1016/j.knsys.2020.105580>
19. Xiao F, Wen J, Pedrycz W (2022) Generalized divergence-based decision making method with an application to pattern classification. *IEEE Trans Knowl Data Eng* :1–1
20. Li Z, Ren T, Ma X, Liu S, Zhang Y, Zhou T (2019) Identifying influential spreaders by gravity model. *Sci Rep* 9(1):1–7. <https://doi.org/10.1038/s41598-019-44930-9>
21. Liu F, Wang Z, Deng Y (2020) GMM: A generalized mechanics model for identifying the importance of nodes in complex networks. *Knowl-Based Syst* 193:105464. <https://doi.org/10.1016/j.knsys.2019.105464>
22. Li H, Shang Q, Deng Y (2021) A generalized gravity model for influential spreaders identification in complex networks. *Chaos, Solitons Fractals* 143:110456. <https://doi.org/10.1016/j.chaos.2020.110456>
23. Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. *J Math Sociol* 2(1):113–120. <https://doi.org/10.1080/0022250X.1972.9989806>
24. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. *Nature Phys* 6(11):888–893. <https://doi.org/10.1038/nphys1746>
25. Zareie A, Sheikhhahmadi A (2018) A hierarchical approach for influential node ranking in complex social networks. *Expert Syst Appl* 93:200–211. <https://doi.org/10.1016/j.eswa.2017.10.018>
26. Maji G, Mandal S, Sen S (2020) A systematic survey on influential spreaders identification in complex networks with a focus on K-shell based techniques. *Expert Syst Appl* 161:113681. <https://doi.org/10.1016/j.eswa.2020.113681>
27. Maji G, Namtirtha A, Dutta A, Malta MC (2020) Influential spreaders identification in complex networks with improved k-shell hybrid method. *Expert Syst Appl* 144:113092. <https://doi.org/10.1016/j.eswa.2019.113092>
28. Maji G (2020) Influential spreaders identification in complex networks with potential edge weight based k-shell degree neighborhood method. *J Comput Sci* 39:101055. <https://doi.org/10.1016/j.jocs.2019.101055>
29. Zareie A, Sheikhhahmadi A (2019) EHC: Extended H-index centrality measure for identification of users' spreading influence in complex networks. *Phys A Stat Mech Appl* 514:141–155. <https://doi.org/10.1016/j.physa.2018.09.064>
30. Freeman LC (1978) Centrality in social networks conceptual clarification. *Soc Netw* 1(3):215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
31. Newman ME (2005) A measure of betweenness centrality based on random walks. *Soc Netw* 27(1):39–54. <https://doi.org/10.1016/j.socnet.2004.11.009>
32. Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. *Soc Netw* 23(3):191–201. [https://doi.org/10.1016/S0378-8733\(01\)00038-7](https://doi.org/10.1016/S0378-8733(01)00038-7)
33. Brin S, Page L (2012) Reprint of: The anatomy of a large-scale hypertextual web search engine. *Comput Netw* 56(18):3825–3833. <https://doi.org/10.1016/j.comnet.2012.10.007>
34. Wang Y, Wang M, Fujita H (2020) Word Sense Disambiguation: A comprehensive knowledge exploitation framework. *Knowl-Based Syst* 190:105030. <https://doi.org/10.1016/j.knsys.2019.105030>
35. Wen T, Deng Y (2020) Identification of influencers in complex networks by local information dimensionality. *Inf Sci* 512:549–562. <https://doi.org/10.1016/j.ins.2019.10.003>
36. Xiao F, Pedrycz W (2022) Negation of the quantum mass function for multisource quantum information fusion with its application to pattern classification. *IEEE Trans Pattern Anal Mach Intell*:1–1. <https://doi.org/10.1109/TPAMI.2022.3167045>
37. Fan W, Xiao F (2022) A complex Jensen–Shannon divergence in complex evidence theory with its application in multi-source information fusion. *Eng Appl Artif Intell* 116:105362. <https://doi.org/10.1016/j.engappai.2022.105362>
38. Deng Y (2022) Random permutation set. *Int J Comput Commun Control* 17(1):4542. <https://doi.org/10.15837/ijccc.2022.1.4542>
39. Zhu C, Xiao F, Cao Z (2022) A generalized Rényi divergence for multi-source information fusion with its application in EEG data analysis. *Inf Sci* 605:225–243. <https://doi.org/10.1016/j.ins.2022.05.012>
40. Xiao F (2022) GEJS: A generalized evidential divergence measure for multisource information fusion. *IEEE Trans Syst Man Cybernet Syst* :1–13
41. Deng Y (2020) Information volume of mass function. *Int J Comput Commun Contr* 15(6):3983. <https://doi.org/10.15837/ijccc.2020.6.3983>
42. Xiao F, Cao Z, Lin C-T (2022) A complex weighted discounting multisource information fusion with its application in pattern classification. *IEEE Trans Knowl Data Eng*:1–16. <https://doi.org/10.1109/TKDE.2022.3206871>
43. Deng Y (2020) Uncertainty measure in evidence theory. *Sci China Inf Sci* 63(11):1–19. <https://doi.org/10.1007/s11432-020-3006-9>
44. Zhang S, Xiao F (2022) A TFN-based uncertainty modeling method in complex evidence theory for decision making. *Inf Sci* 619:193–207. <https://doi.org/10.1016/j.ins.2022.11.014>
45. Xiao F (2022) Generalized quantum evidence theory. *Appl Intell*:1–16. <https://doi.org/10.1007/s10489-022-04181-0>
46. Zheng L, Xiao F (2022) Complex interval number-based uncertainty modeling method with its application in decision fusion. *Int J Intell Syst*:1–18. <https://doi.org/10.1002/int.23070>
47. Qiang C, Deng Y, Cheong KH (2022) Information fractal dimension of mass function, vol 30. <https://doi.org/10.1142/S0218348X22501109>
48. Wen T, Jiang W (2019) Identifying influential nodes based on fuzzy local dimension in complex networks. *Chaos, Solitons Fractals* 119:332–342. <https://doi.org/10.1016/j.chaos.2019.01.011>
49. Floyd RW (1962) Algorithm 97: shortest path. *Commun ACM* 5(6):345. <https://doi.org/10.1145/367766.368168>
50. Gleiser PM, Danon L (2003) Community structure in jazz. *Adv Compl Syst* 6(04):565–573. <https://doi.org/10.1142/S0219525903001067>
51. Newman ME (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E* 74(3):036104. <https://doi.org/10.1103/PhysRevE.74.036104>
52. Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: Densification and shrinking diameters. *ACM Trans Knowl Disc Data (TKDD)* 1(1):2–es. <https://doi.org/10.1145/1217299.1217301>

53. Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A (2003) Self-similar community structure in a network of human interactions. *Phys Rev E* 68(6):065103. <https://doi.org/10.1103/PhysRevE.68.065103>
54. Yin H, Benson AR, Leskovec J, Gleich DF (2017) Local higher-order graph clustering. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp 555–564. <https://doi.org/10.1145/3097983.3098069>
55. Adamic LA, Glance N (2005) The political blogosphere and the 2004 US election: divided they blog. In: Proceedings of the 3rd international workshop on link discovery, pp 36–43. <https://doi.org/10.1145/1134271.1134277>
56. McAuley J, Leskovec J (2014) Discovering social circles in ego networks. *ACM Trans Knowl Disc Data (TKDD)* 8(1):1–28. <https://doi.org/10.1145/2556612>
57. Leskovec J, Huttenlocher D, Kleinberg J (2010) Predicting positive and negative links in online social networks. In: Proceedings of the 19th international conference on world wide web, pp 641–650. <https://doi.org/10.1145/1772690.1772756>
58. Colizza V, Pastor-Satorras R, Vespignani A (2007) Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Phys* 3(4):276–282. <https://doi.org/10.1038/nphys560>
59. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442. <https://doi.org/10.1038/30918>
60. Spring N, Mahajan R, Wetherall D, Anderson T (2004) Measuring ISP topologies with Rocketfuel. *IEEE/ACM Trans Netw* 12(1):2–16. <https://doi.org/10.1109/TNET.2003.822655>
61. Yang M, Chen G, Fu X (2011) A modified SIS model with an infective medium on complex networks and its global stability. *Phys A Stat Mech Appl* 390(12):2408–2413. <https://doi.org/10.1016/j.physa.2011.02.007>
62. Castellano C, Pastor-Satorras R (2010) Thresholds for epidemic spreading in networks. *Phys Rev Lett* 105(21):218701. <https://doi.org/10.1103/PhysRevLett.105.218701>
63. Kendall MG (1938) A new measure of rank correlation. *Biometrika* 30(1/2):81–93. <https://doi.org/10.1093/biomet/30.1-2.81>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.