SCIENTIFIC REPORTS
natureresearch

OPEN

Identifying vital nodes in complex networks by adjacency information entropy

Xiang Xu*, Cheng Zhu*, Qingyong Wang, Xianqiang Zhu & Yun Zhou

Identifying the vital nodes in networks is of great significance for understanding the function of nodes and the nature of networks. Many centrality indices, such as betweenness centrality (BC), eccentricity centrality (EC), closeness centricity (CC), structural holes (SH), degree centrality (DC), PageRank (PR) and eigenvector centrality (VC), have been proposed to identify the influential nodes of networks. However, some of these indices have limited application scopes. EC and CC are generally only applicable to undirected networks, while PR and VC are generally used for directed networks. To design a more applicable centrality measure, two vital node identification algorithms based on node adjacency information entropy are proposed in this paper. To validate the effectiveness and applicability of the proposed algorithms, contrast experiments are conducted with the BC, EC, CC, SH, DC, PR and VC indices in different kinds of networks. The results show that the index in this paper has a high correlation with the local metric DC, and it also has a certain correlation with the PR and VC indices for directed networks. In addition, the experimental results indicate that our algorithms can effectively identify the vital nodes in different networks.

The vital nodes in networks are the nodes that have great impacts on the network structure and function[1]. Previous studies have described many centralities that can rank the nodes in networks, such as degree centrality[2], eccentricity[3], closeness centricity[4], betweenness centrality[5–7], eigenvector centrality[8] and PageRank[9]. Identifying the influential nodes in networks is not only of theoretical significance but also of practical value. For example, identifying the important junctions in traffic networks can prevent the paralysis of traffic networks caused by traffic congestion. Locking key sources in virus transmission networks can significantly reduce the speed and scope of virus transmission. These examples and others are all related to identifying the vital nodes in networks. The paper of Gino *et al.* applied the optimal percolation theory to predict the influential nodes in memory networks[10].

Considering that the local metrics have lower computational complexity and the global metrics have higher computational accuracy, in recent work, many vital node identification methods that consider both local and global metrics have been proposed. A semi-local metric that balances the accuracy and efficiency was proposed by Chen *et al.*[11]. Another neighbourhood centrality that takes into account the importance of a node and its neighbours' was proposed[12]. In the paper by Yu *et al.*[13], an improved method called improved structural holes (ISH) that identifies the key nodes in complex networks was proposed; unlike the eccentricity and betweenness centrality, this method can be applied to large-scale and disconnected networks. Zhang *et al.*[14] presented an effective method named VoteRank to identify a set of dispersive spreaders with the best spreading ability. By considering the propagation probability, Ma *et al.*[15] proposed a new algorithm named hybrid degree centrality (HC) to improve the local metrics and combined it with degree centrality. Lü *et al.*[16] gave a complete overview of the vital node identification methods in recent years.

In addition to the above perspectives, many other related references use network dynamics to study the importance of nodes in networks. Lü *et al.*[17] devised an adaptive and parameter-free algorithm, the LeaderRank, to measure the influence of users in social networks, and the experimental results show that the algorithm is more efficient than PageRank and more robust to noisy data. Min[18] proposed a method using a message-passing approach for identifying the most influential spreaders in networks and found that the method can be easily applied to unweighted and weighted networks. Liu *et al.*[19] presented dynamics-sensitive (DS) centrality for locating influential nodes by combining the topological and dynamic characteristics of the networks. Zhang *et al.*[20] designed a multiscale node-importance method to measure the importance of nodes in the process of network dynamics according to different network scales. Many related studies only identify the vital nodes for a certain

Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha, 410072, China. *email: xuxiang19@nudt.edu.cn; zhucheng@nudt.edu.cn

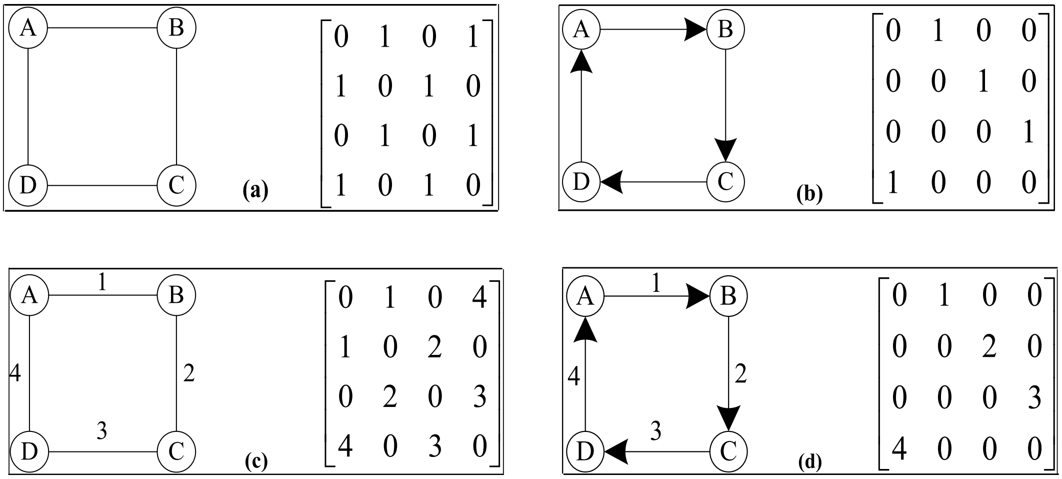

Figure 1. Four different types of networks and their corresponding adjacency matrices. (**a**) An unweighted-undirected network and its adjacency matrix. (**b**) An unweighted-directed network and its adjacency matrix. (**c**) A weighted-undirected network and its adjacency matrix. (**d**) A weighted-directed network and its adjacency matrix.

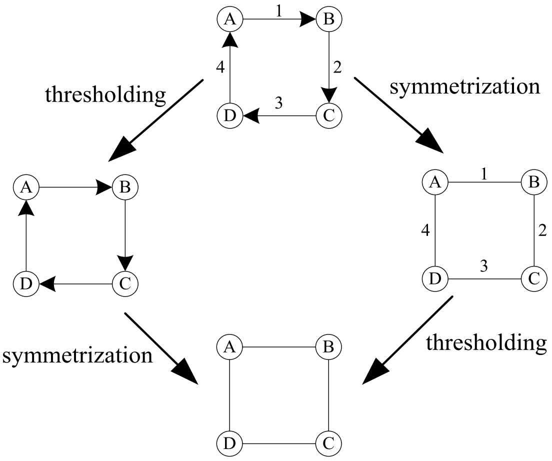

Figure 2. Relationships among the four different types of networks. Directed networks can obtain their corresponding undirected networks via symmetrisation, and weighted networks can obtain their corresponding unweighted networks via thresholding.

type of network, such as refs. [21–23] that only study node identification for weighted networks, Chen *et al.*[24] who proposed an identification method for directed networks, reference[25] that mined the vital nodes in directed weighted complex networks, and Edgar *et al.*[26] who used Kuramoto and Ising dynamics to study the central role of peripheral nodes in directed networks. The last paper argued that a large key component does not uniquely ensure the emergence of collective phenomena, as it does for undirected networks.

To identify the vital nodes for different types (unweighted-undirected, unweighted-directed, weighted-undirected and weighted-directed) of networks, we propose an adjacency information entropy method to identify the vital nodes in different networks by considering the weights and directions of the edges in networks. For weighted networks, the node strength is used instead of the node degree. For directed networks, in order to refine the influence of the out-degree and in-degree on the node importance, we set the influence coefficient θ of the in-degree value. By adjusting the size of θ, we can control the different influences of the out-degree and in-degree on nodes.

The rest of this paper is organised as follows. In section 2, we provide detailed representations of the different types of networks. In section 3, two vital node identification algorithms and three related definitions are proposed. In section 4, four empirical experiments on the independent parts, largest components, network efficiency and correlations analysis are carried out, and the experimental results are compared and explained. Finally, the conclusion and future works are presented in section 5.

Representations

In this paper, we study vital node identification in four different types of networks, namely, unweighted-undirected networks, unweighted-directed networks, weighted-undirected networks and weighted-directed networks. Obviously, the representations of the different networks and the calculations of the related metrics in the networks are different.

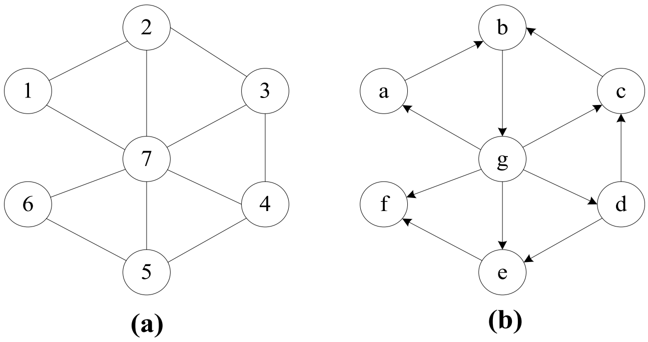

Figure 3. Two example networks with 7 nodes and 11 edges. (**a**) An unweighted-undirected network. (**b**) An unweighted-directed network.

UUNs	***n***	***m***	$<\boldsymbol{k}>$	$<\boldsymbol{d}>$	***C***
Astro	14845	239304	16.12	4.798	0.715
CA	8638	49612	5.743	5.945	0.580
Facebook	4039	88234	43.69	3.693	0.617
Hamster	2000	32194	16.09	3.589	0.573
UDNs	***n***	***m***	$<\boldsymbol{k}>$	$<\boldsymbol{d}>$	***C***
Email	1133	5451	4.811	3.715	0.110
PGP	10680	24340	2.279	4.050	0.133
Router	5022	6258	1.246	3.973	0.006
Wiki-Vote	7115	103689	14.57	3.341	0.081
WUNs	***n***	***m***	$<\boldsymbol{k}>$	$<\boldsymbol{d}>$	***C***
Astro	14845	239304	1256.6	4.798	0.715
CA	8638	49612	448.78	5.945	0.580
Facebook	4039	88234	1113.7	3.693	0.617
Hamster	2000	32194	1257.2	3.589	0.573
WDNs	***n***	***m***	$<\boldsymbol{k}>$	$<\boldsymbol{d}>$	***C***
Email	1133	5451	4.811	3.715	0.110
P2P	6301	20777	29.663	6.632	0.005
PHD	1025	1043	8.956	3.429	0.002
Router	5022	6258	1.246	3.973	0.006

Table 1. The statistical properties of the four kinds of complex networks, where ***n*** and ***m*** are the total numbers of nodes and edges, respectively. $<\boldsymbol{k}>$ and $<\boldsymbol{d}>$ denote the average degree and the average distance respectively, and ***C*** denotes the clustering coefficient.

Four different types of networks.

Usually, an unweighted network is represented by $G = (V, E)$, where $V = \{v_1, v_2, \cdots, v_n\}$ and $|V|$ is the number of nodes in the network. $E = \{e_1, e_2, \cdots, e_m\}$ and $|E|$ is the number of edges in the network. An adjacency matrix is used to represent the connections between the nodes in the network, and the topology of the network can be obtained by using the adjacency matrix. In Fig. 1(a), the left figure is an unweighted-undirected network and the right figure is its corresponding adjacency matrix. It is obvious that the adjacency matrix of undirected network is a symmetric matrix.

The degree of nodes in unweighted-undirected networks can be calculated by $k_i = \sum_{j=1}^{m} a_{ij}$, where j is the neighbour of node i and m is the number of neighbours of node i. $a_{ij} = 1$ if there is an edge between node i and node j, and otherwise it is 0.

Unlike unweighted-undirected networks, the edges between the nodes in unweighted-directed networks have directions. The asymmetry of the adjacency matrix can reflect the directions of the edges in networks. We can see that the matrix in Fig. 1(b) is different from the matrix in Fig. 1(a). There are two kinds of degrees of nodes in directed networks, namely, in-degree and out-degree. In directed networks, the in-degree of a node is the number of edges from its neighbours that point to it, and the out-degree of a node is the number of edges of the node that point to its neighbours. These two kinds of degrees can be calculated by Eqs. 1 and 2, respectively.

$$k_i^{in} = \begin{cases} \sum_{j \in \Gamma_i} a_{ji} & \text{if the network is unweighted} \\ \sum_{j \in \Gamma_i} w_{ji} & \text{if the network is weighted} \end{cases} \tag{1}$$

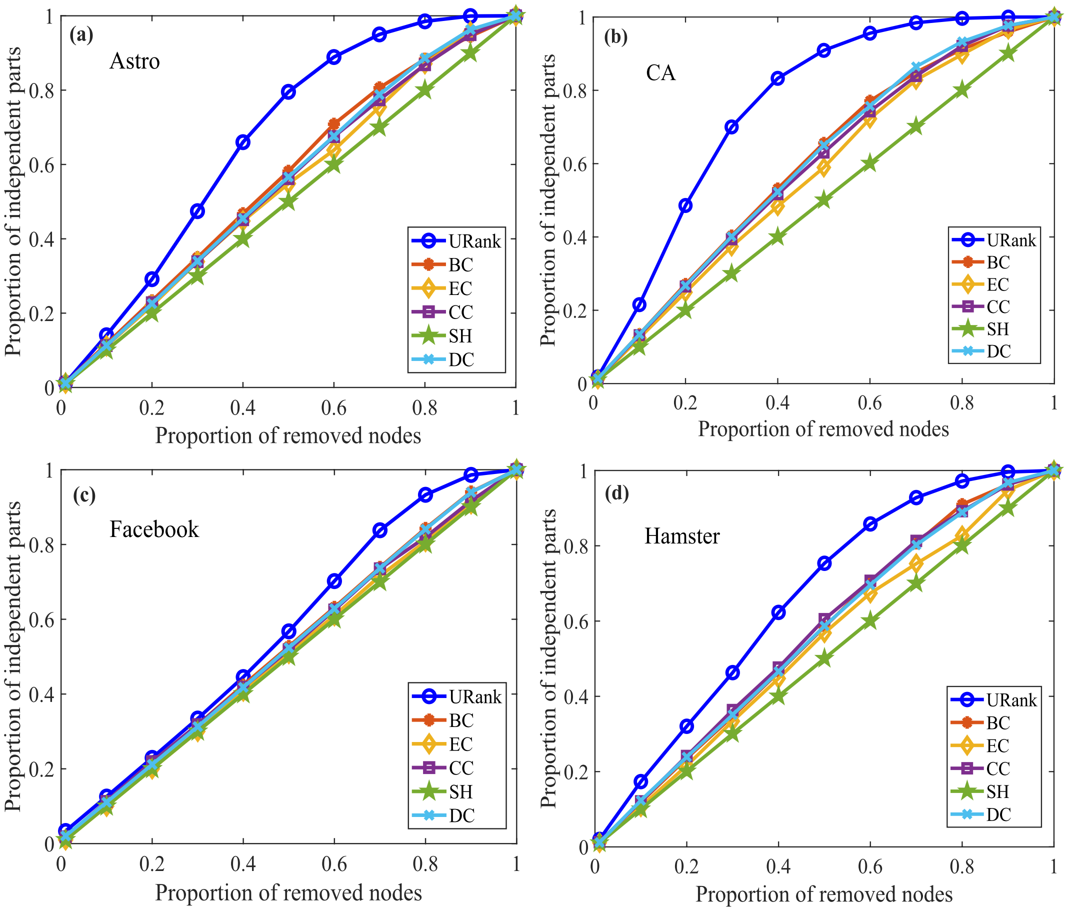

Figure 4. Independent parts experiments for unweighted-undirected networks. In (**a**), when the proportion of removed nodes reaches approximately 10%, the performance of URank is better than the other metrics; and when more than 20% of the nodes are removed, URank performs better and better until approximately 60% of the nodes are removed. Panel (**b**) clearly shows the performance of URank compared with the other metrics. In (**c**), it can be seen that there is little difference among the different metrics when removing the first 40% of the nodes. Panel (**d**) shows that the performance of URank is more uniform.

$$k_i^{out} = \begin{cases} \sum_{j \in \Gamma_i} a_{ij} & \text{if the network is unweighted} \\ \sum_{j \in \Gamma_i} w_{ij} & \text{if the network is weighted} \end{cases} \tag{2}$$

Generally, the calculation of the degree in directed networks adds the in-degree to the out-degree. Here, we consider that the in-degree and out-degree of nodes have different effects on nodes[27]. Then, the degree of nodes in directed networks can be calculated by Eq. 3, where θ is the influence coefficient of the nodes' in-degree, and in this paper, we set $\theta = 0.75$.

$$k_i^{unweighted} = \theta k_i^{in} + (1 - \theta) k_i^{out} = \theta \sum_{j=1}^{m} a_{ji} + (1 - \theta) \sum_{j=1}^{m} a_{ij} \tag{3}$$

A weighted network can be represented by $G = (V, E, W)$, where W is the adjacency weighted matrix of the network. The weights of the connected edges in weighted networks are not only 0 or 1, and edges' weights can reflect the strength of the relationships between nodes. Figure 1(c) presents a weighted-undirected network and the corresponding adjacency matrix. The degree of nodes in weighted-undirected networks can be obtained by $k_i = \sum_{j=1}^{m} w_{ij}$, where w_{ij} is the weight of the edge between node i and node j.

Weighted-directed networks are the most complicated of the four types of networks. Figure 1(d) shows a simple weighted-directed network and its adjacency weighted matrix. According to the above degree calculation method for weighted networks and directed networks, naturally, the degree of nodes in weighted-directed networks can be obtained by Eq. 4. Figure 2 illustrates the relationships among the four different types of networks.

$$k_i^{weighted} = \theta k_i^{in} + (1 - \theta) k_i^{out} = \theta \sum_{j=1}^{m} w_{ji} + (1 - \theta) \sum_{j=1}^{m} w_{ij} \tag{4}$$

Methods

Related definitions.

To identify the vital nodes in different types of networks, we propose three definitions as follows.

Definition 1. Adjacency degree A_i. We define the adjacency degree of nodes in undirected networks by considering its nearest neighbours as $A_i = \sum_{j\in\Gamma_i} k_j$, where *j* is the neighbour of node *i*, Γ_i is the set of neighbours of node *i*, and k_j is the degree of node *j*. For example, in Fig. 3(a), $A_1 = k_2 + k_7 = 3 + 6 = 9$. In directed networks, the adjacency degree of nodes is defined as follows (Eq. 5), where $k_{j_{in}}$ is the number of edges that point to node *j* from node *i*, and $k_{j_{out}}$ is the number of edges from node *j* that point to node *i*. For example, in Fig. 3(b), $A_b = \theta(k_a + k_c) + (1 - \theta)k_g = \theta * (1 + 1.75) + (1 - \theta) * 2 = 2.5625$.

$$A_i = \theta\sum_{j\in\Gamma_i} k_{j_{in}} + (1 - \theta)\sum_{j\in\Gamma_i} k_{j_{out}} \tag{5}$$

Definition 2. Selection probability P_{i_j}. We define the selection probability of node *i* in the network by considering the probability that it will be selected by its neighbour *j*, and the calculation formula is Eq. 6.

Taking from the idea from information theory, a certain node in the network is taken as the information source point, and its neighbouring nodes are taken as the target points. In the process of information transmission or disease transmission, the information source point and infected person will select the target point among its neighbouring nodes for information transmission or disease infection. The probability that the target nodes are selected is called the selection probability. This definition considers the importance of the selected nodes, that is, the influence of the degrees of the selected node in the selection process.

$$P_{i_j} = k_i / A_j,\ (j \in \Gamma_i) \tag{6}$$

For example, in Fig. 3(a) $P_{1_2} = k_1/A_2 = k_1/(k_1 + k_3 + k_7) = 2/(2 + 3 + 6) \approx 0.23$. Similarly, $P_{1_7} = k_1/A_7 = k_1/(k_1 + k_2 + k_3 + k_4 + k_5 + k_6) = 2/(2 + 3 + 3 + 3 + 3 + 2) = 0.125$.

Definition 3. Adjacency information entropy E_i. We define the adjacency information entropy of nodes in undirected networks as Eq. 7 and that in directed networks as Eq. 8.

$$E_i = -\sum_{j\in\Gamma_i}\left(P_{i_j} log_2 P_{i_j}\right) \tag{7}$$

$$E_i = \sum_{j\in\Gamma_i}\left|\left(-P_{i_j} log_2 P_{i_j}\right)\right| \tag{8}$$

Vital node identification algorithms.

According to the characteristics of the four different types of networks, the proposed algorithms in this paper can be applied to different networks. Before the algorithms can be applied, we need to obtain the adjacency matrix ***A*** or the adjacency weighted matrix ***W***of the network. From the above definitions, we can rank the nodes in the network by the value of the node's adjacency information entropy (E_i), and the specific algorithms step are as follows.

Algorithm 1. Calculate F_i for unweighted networks.

```
Input: The adjacency matrix $\boldsymbol{A}$ of corresponding networks.
Output: Adjacency information entropy value $E_i$ of nodes in corresponding networks.
for $i \leftarrow 1$ to $N$ do
  if $A_{ij} = A_{ji}$ then
    $k_i \leftarrow \sum_{j=1}^{m} a_{ij}$
    $A_i \leftarrow \sum_{j\in\Gamma_i} k_j$
    $P_{i_j} \leftarrow k_i/A_j$
    $E_i \leftarrow -\sum_{j\in\Gamma_i}(P_{i_j} log_2 P_{i_j})$
  else
    $k_i \leftarrow \theta\sum_{j=1}^{m} a_{ji} + (1-\theta)\sum_{j=1}^{m} a_{ij}$
    $A_i \leftarrow \theta\sum_{j\in\Gamma_i} k_{j_{in}} + (1-\theta)\sum_{j\in\Gamma_i} k_{j_{out}}$
    $P_{i_j} \leftarrow k_i/A_j$
    $E_i \leftarrow \sum_{j\in\Gamma_i}|(-P_{i_j} log_2 P_{i_j})|$
  end if
end for
return $E_i$
```

Algorithm 2. Calculate F_i for weighted networks.

```
Input: The weighted matrix $\boldsymbol{W}$ of corresponding networks.
Output: Adjacency information entropy value $E_i$ of nodes in corresponding networks.
for $i \leftarrow 1$ to $N$ do
  if $W_{ij} = W_{ji}$ then
    $k_i \leftarrow \sum_{j=1}^{m} w_{ij}$
    $A_i \leftarrow \sum_{j\in\Gamma_i} k_j$
    $P_{i_j} \leftarrow k_i/A_j$
    $E_i \leftarrow -\sum_{j\in\Gamma_i}(P_{i_j} log_2 P_{i_j})$
  else
    $k_i \leftarrow \theta\sum_{j=1}^{m} w_{ji} + (1-\theta)\sum_{j=1}^{m} w_{ij}$
    $A_i \leftarrow \theta\sum_{j\in\Gamma_i} k_{j_{in}} + (1-\theta)\sum_{j\in\Gamma_i} k_{j_{out}}$
    $P_{i_j} \leftarrow k_i/A_j$
    $E_i \leftarrow \sum_{j\in\Gamma_i}|(-P_{i_j} log_2 P_{i_j})|$
  end if
end for
return $E_i$
```

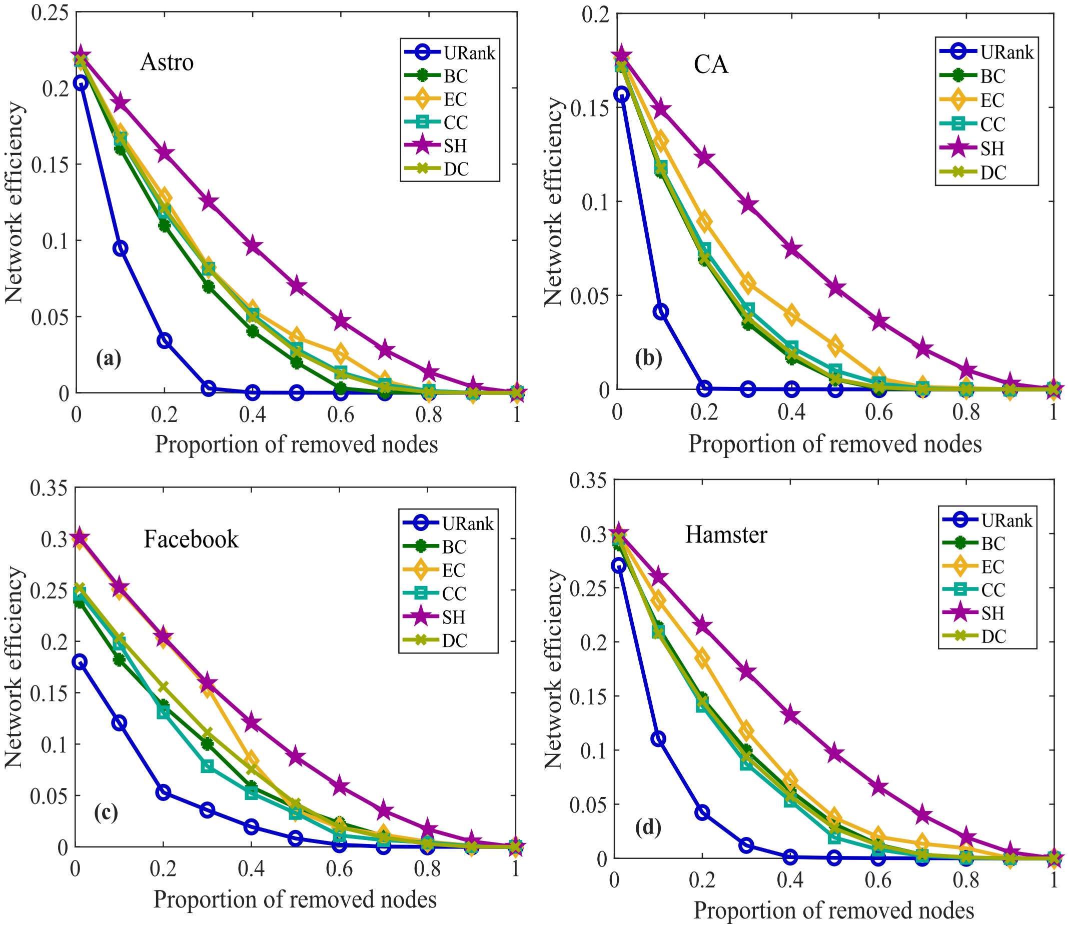

Figure 5. Network efficiency experiments for undirected networks. In (**a**), as the proportion of removed nodes increases, the network efficiency gradually declines after removing the nodes ranked by different metrics. Node removal ranked by URank makes the network efficiency decrease the fastest, and the network efficiency corresponding to URank is the smallest when the same proportion of nodes are removed. In (**b**), after removing approximately 20% of the nodes ranked by the algorithm in this paper, the network efficiency is reduced to the lowest, while the other metrics need to remove approximately 70% of the nodes. Panel (**c**) shows that the performance of our metric is more uniform. Panel (**d**) clearly shows the performance of URank compared with the other metrics.

Results and Discussion

To verify the accuracy and applicability of our proposed algorithms, four different kinds of networks are employed, which include (1) unweighted-undirected networks (UUNs), (2) unweighted-directed networks (UDNs), (3) weighted-undirected networks (WUNs), and (4) weighted-directed networks (WDNs). The statistical properties of the studied networks are listed in Table 1. With respect to the unweighted-undirected networks, the Astro network is a collaboration network of astrophysics scientists[28]; the CA network is a large connected component of the arXiv collaboration network in high-energy physics theory[29]; the Facebook network is an anonymised social networks with 4039 users, where the data can be downloaded in http://snap.stanford.edu/data/; and the Hamster network is a friendship and family connections network among website users[30]. With respect to the unweighted-directed networks, the Email network includes 1133 email users of the University at Rovira i Virgili, URV[31]; the PGP network is a communication network[32]; Router is a topological network of the Internet[33]; and the Wiki-Vote network is a who-votes-on-whom network from Wikipedia, where the data can be downloades from http://snap.stanford.edu/data/. With respect to the weighted-directed networks, the data of the P2P and PHD networks can be obtained at http://vlado.fmf.uni-lj.si/pub/networks/data/.

We will verify the accuracy of our algorithms by computing the proportion of independent parts of networks by removing the different proportions of nodes. Obviously, the larger the proportion of independent parts is, the more seriously the network is destroyed, and the higher the identification accuracy of vital nodes is. For undirected networks, we selected five other centralities as benchmark indices, namely, betweenness centrality (BC), eccentricity centrality (EC), closeness centrality (CC), structural holes centrality[34] (SH) and degree centrality (DC). For simplicity, we call our algorithm for unweighted-undirected networks URank and that for weighted-undirected networks WRank. The X-axis is the different proportions of removed nodes. The Y-axis is the proportions of independent parts of corresponding networks. The results of the different centralities after

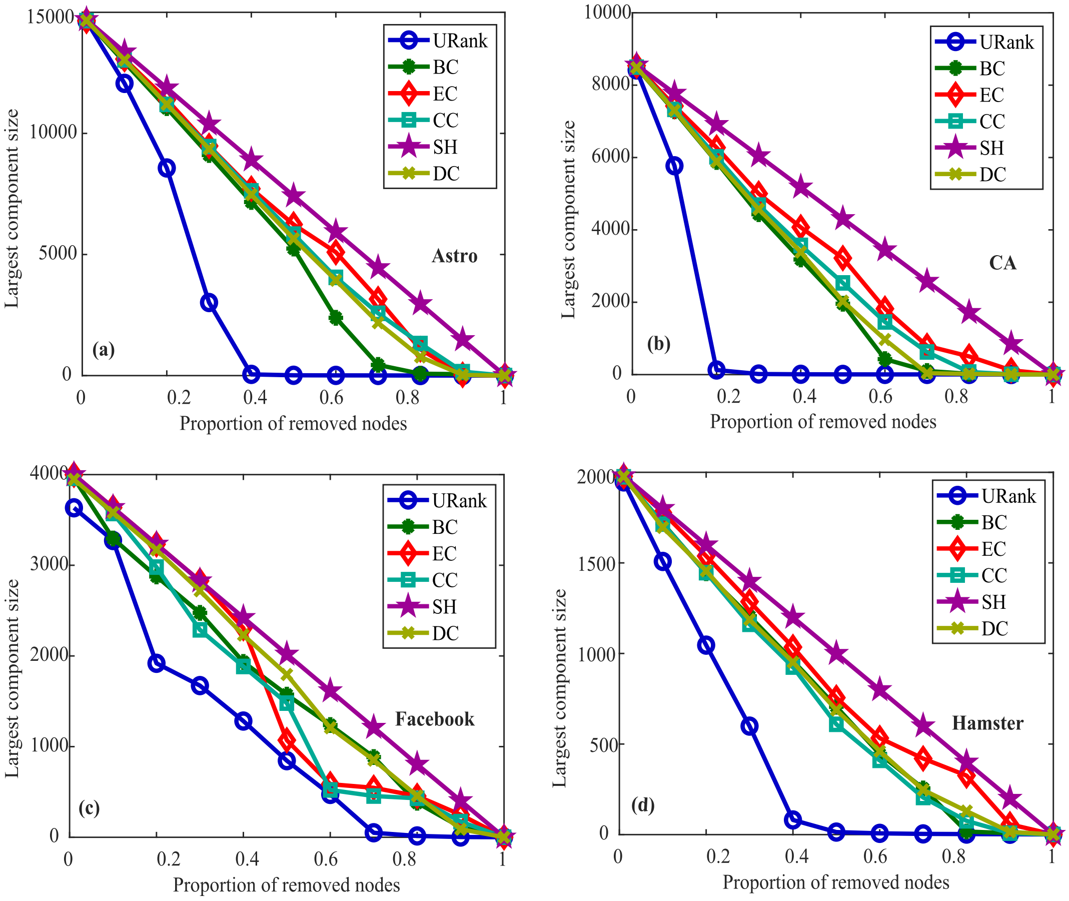

Figure 6. Largest component experiments for unweighted-undirected networks. In (**a**), after removing approximately 40% of the nodes ranked by our algorithm, the largest component size is reduced to the lowest, while the other metrics need to remove approximately 80% or 90% of the nodes. Panel (b)clearly shows that the performance of URank compared with other metrics. In (**c**), the figure shows the performance of our metric is more uniform, and there is little difference in the removal ratio of nodes when each metric makes the largest component size reach the lowest. In panel (d), when removing the first 40% of the nodes, our metric makes the largest component size decline the fastest until it reaches the minimum.

removing different proportions of ranked nodes in the four different unweighted-undirected networks are shown in Fig. 4. From Figure 4, it is clear that our algorithm is more significant. Similarly, Supplementary Fig. S1 shows the experimental results for weighted-undirected networks. In addition, according to the importance of the nodes, we also remove the different proportions of nodes from high to low to test the efficiency of the undirected networks. As is well known, the higher the network efficiency is, the smaller the average distance between the nodes in the network is. If the removed nodes cause the network efficiency to decline more, the impact of the removed nodes on the network is greater, the removed nodes are more important. Figure 5 shows the experimental results of the network efficiency curves of undirected networks after removing different proportions of nodes.

For directed networks, we also selected five other centralities as benchmark indices, namely, PageRank centrality (PR), eigenvector centrality (VC), eccentricity centrality (EC), closeness centrality (CC) and degree centrality (DC). Similarly, for simplicity, we call our algorithm for unweighted-directed networks DRank and that for weighted-directed networks WDRank. Supplementary Figs. S2 and S3 show the independent parts experimental results in unweighted-directed networks and weighted-directed networks, respectively.

To further prove the effectiveness and applicability of the proposed algorithms, we implemented the largest component experiments using the four different types of networks. When some nodes in the network are deleted according to the importance of the nodes, different sized components will be formed. If the size of the component is smaller, the removed nodes are more destructive to the original network. The largest component experiments can illustrate the accuracy of the vital node identification algorithms from another perspective. The X-axis is the different proportions of removed nodes. The Y-axis is the largest component sizes of the different networks when the corresponding proportion nodes were removed. Figure 6 shows the experimental results for unweighted-undirected networks. Figure 6, shows that the URank algorithm performs well for most networks. The results of the same largest component experiments for unweighted-directed networks, weighted-undirected networks and weighted-directed networks are shown in Supplementary Figs. S5–S6, respectively.

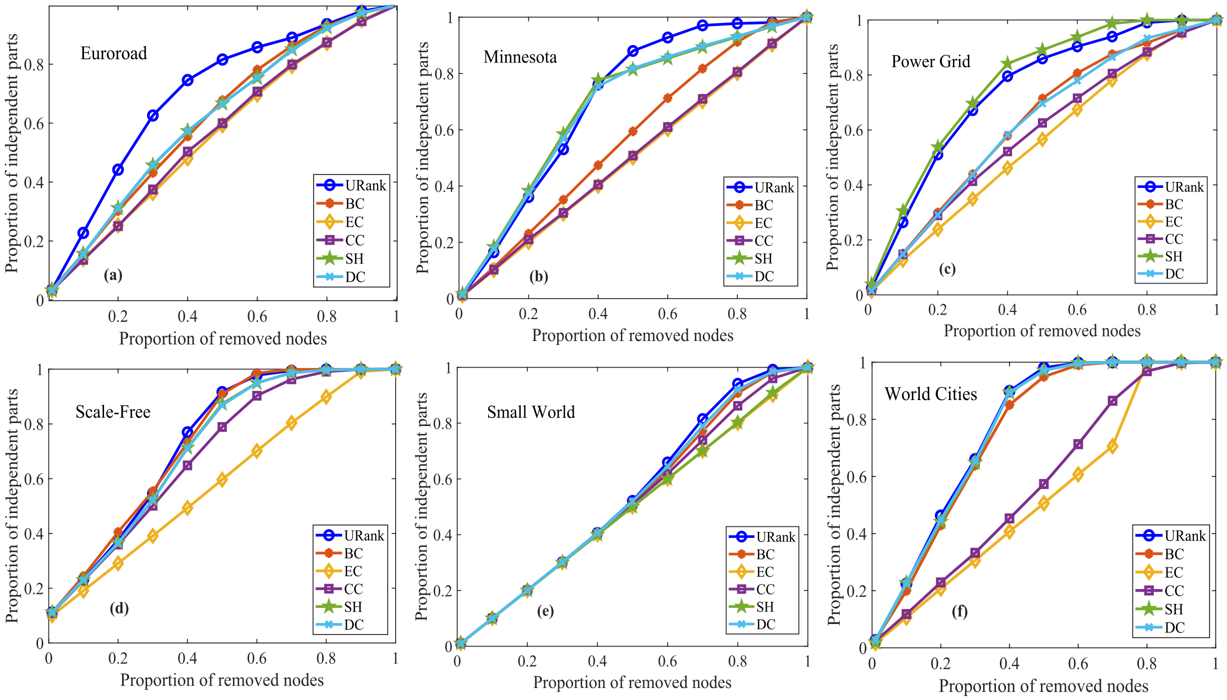

Figure 7. Independent parts experiments for spatial networks and classical networks. In (**a**), as the number of removed nodes increase, the proportion of parts corresponding to URank is significantly more than those of the other metrics. From (**b**,**d**,**e**), we can see that the performance of URank is less obvious than that in (**a**). In (**c**), URank performs better than the other metrics except the SH metric. In (**f**), the URank performs as well as DC, and it is better than the other metrics.

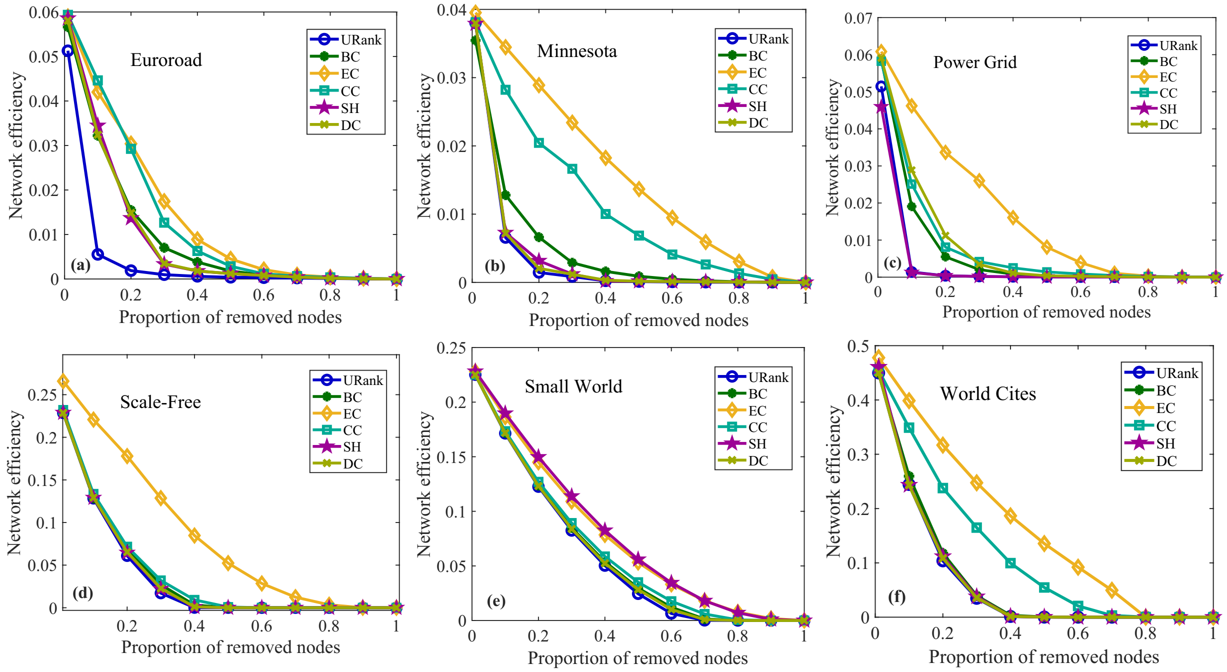

Figure 8. Network efficiency experiments for spatial networks and classical networks. In (**a**), compared with the other metrics, removing nodes ranked by URank makes the network efficiency decline the fastest, and the extent of the decline is also the largest. However, in (**b**–**f**), the advantage of URank is not obvious and, generally speaking, the performance of the metric in this paper is better than those of other metrics.

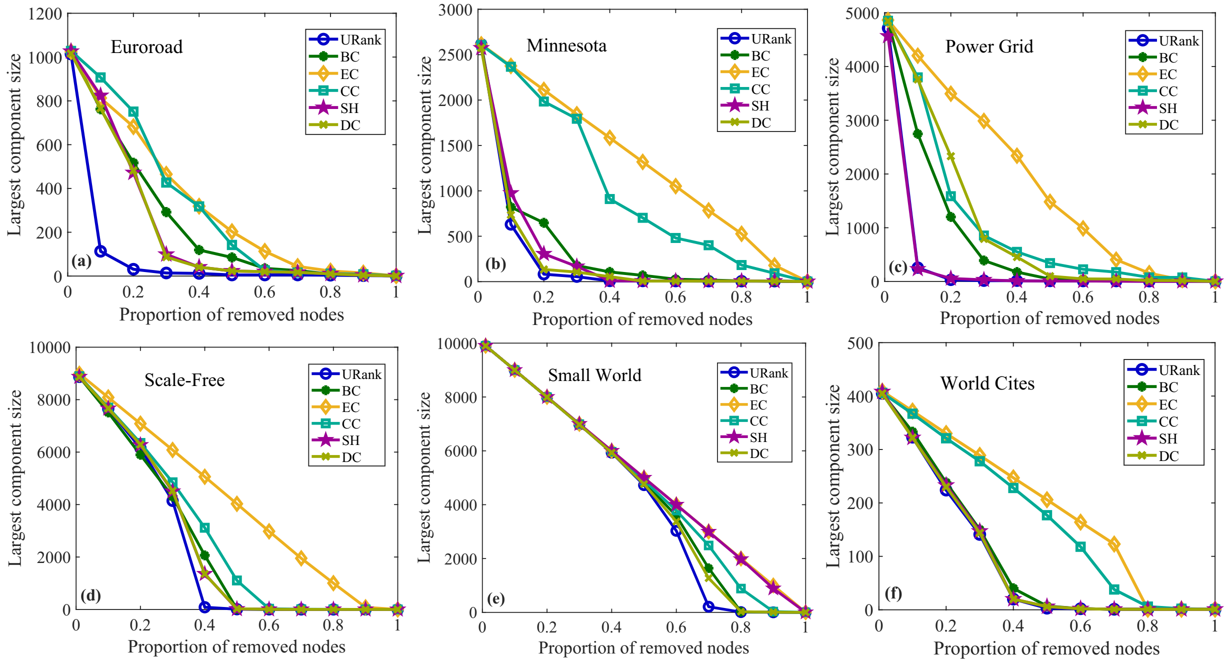

Figure 9. Largest component experiments for spatial networks and classical networks. In (**a**), we can see that when the nodes in the network are removed according to URank, removing approximately 10% of the nodes reduced the largest component size of the network by approximately 90%. If the nodes are removed after being ranked by other metrics, more than 30% of the nodes need to be removed to achieve the same effect. In (**b**), the performance of our metric is slightly better than other metrics. In (**c**,**f**), URank and SH perform equally well. In (**d**,**e**), the performance of URank gradually gets better as the node removal ratio increases until the largest component size of the network reaches the minimum.

Other networks	n	m	$<k>$	$<d>$	C
Euroroad	1174	1417	2.414	18.371	0.020
Minnesota	2642	3303	2.500	35.349	0.017
Power Grid	4941	6594	2.669	18.989	0.107
Scale-Free	10000	187396	18.740	3.223	0.042
Small World	10000	100000	10.000	4.443	0.090
World Cites	415	7518	36.231	2.238	0.003

Table 2. The statistical properties of spatial networks and classical networks, where $\boldsymbol{n}$ and $\boldsymbol{m}$ are the total numbers of nodes and edges, respectively. $<\boldsymbol{k}>$ and $<\boldsymbol{d}>$ denote the average degree and the average distance, respectively, and $\boldsymbol{C}$ denotes the clustering coefficient.

To verify the applicability of our algorithms to other kinds of networks, we further carry out the three verification experiments described above using spatial networks and classical networks, such as a small world network and scale-free network, respectively. Figures 7, 8 and 9 present the results of the independent parts experiments, the network efficiency experiments and the largest component experiments, respectively. The corresponding statistical properties of the spatial networks and classical networks are listed in Table 2. Euroroad and Minnesota are road networks, and the data can be downloaded from http://networkrepository.com/road.php. Power Grid[35] contains an undirected unweighted representation of the topology of the Western States Power Grid of the United States, which was compiled by Duncan Watts and Steven Strogatz. The data are downloaded from the web site of Prof. Duncan Watts at Columbia University, http://cdg.columbia.edu/cdg/datasets. The Scale-Free and Small World networks are generated by the Pajek software. World Cites is a network of 415 cities, and the data can be obtained from http://www-personal.umich.edu/mejn/netdata/.

To investigate the relations between our algorithms and other centralities in different networks, we conducted correlation analysis experiments. We use the Kendall's Tau to describe the relationship between different centralities. The relevant definitions are as follows[36].

Assuming that two random variables are X and Y (they can also be regarded as two sets), their number of elements is N, where X_i and Y_i represent the i-th element of each random variable, respectively. The corresponding elements in X and Y form an element pair set XY, which contains the elements $(X_i, Y_i)(1 \le i \le N)$. When $X_i > X_j$ and $Y_i > Y_j$ or $X_i < X_j$ and $Y_i < Y_j$, these two elements are considered to be concordant. When $X_i > X_j$ and $Y_i < Y_j$ or $X_i < X_j$ and $Y_i > Y_j$, these two elements are considered to be discordant. When $X_i = X_j$ or $Y_i = Y_j$, the two elements are neither concordant nor discordant. Kendall's Tau is defined as.

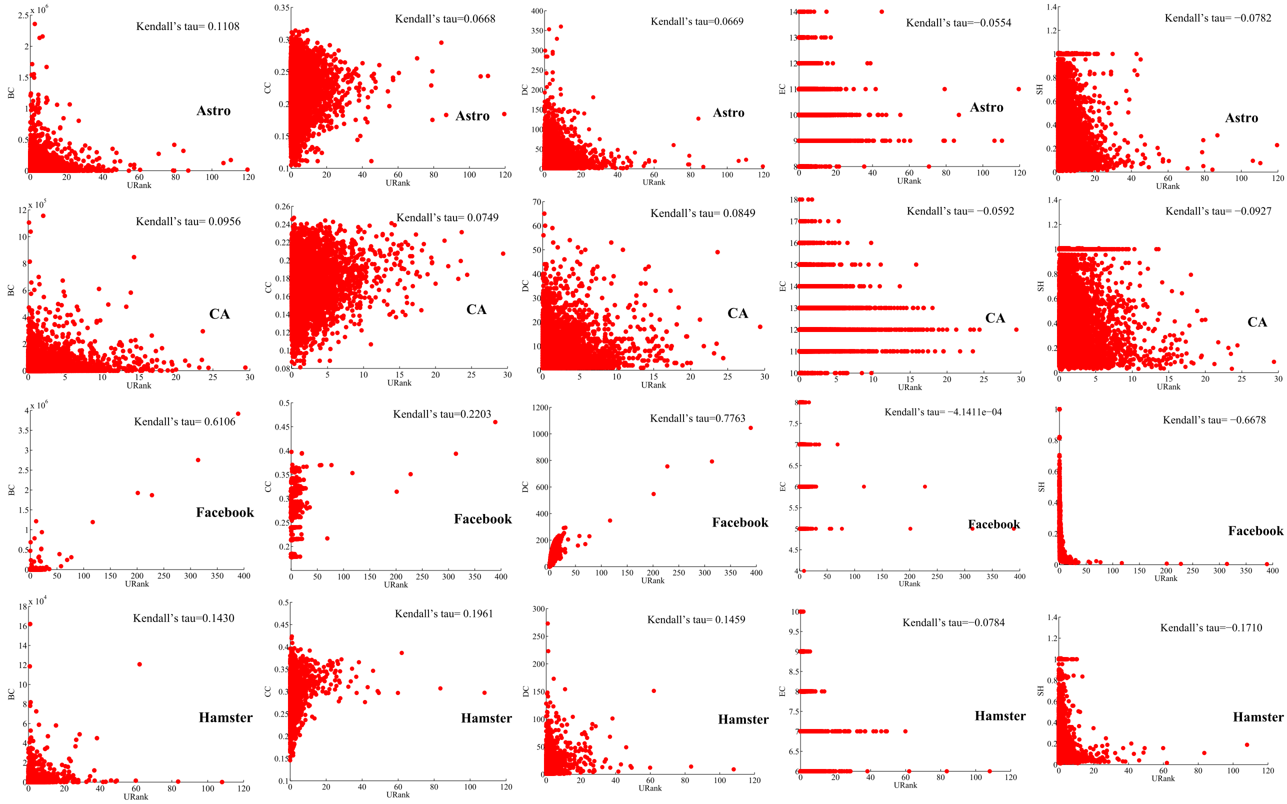

Figure 10. Correlation analysis experiments in unweighted-undirected networks.

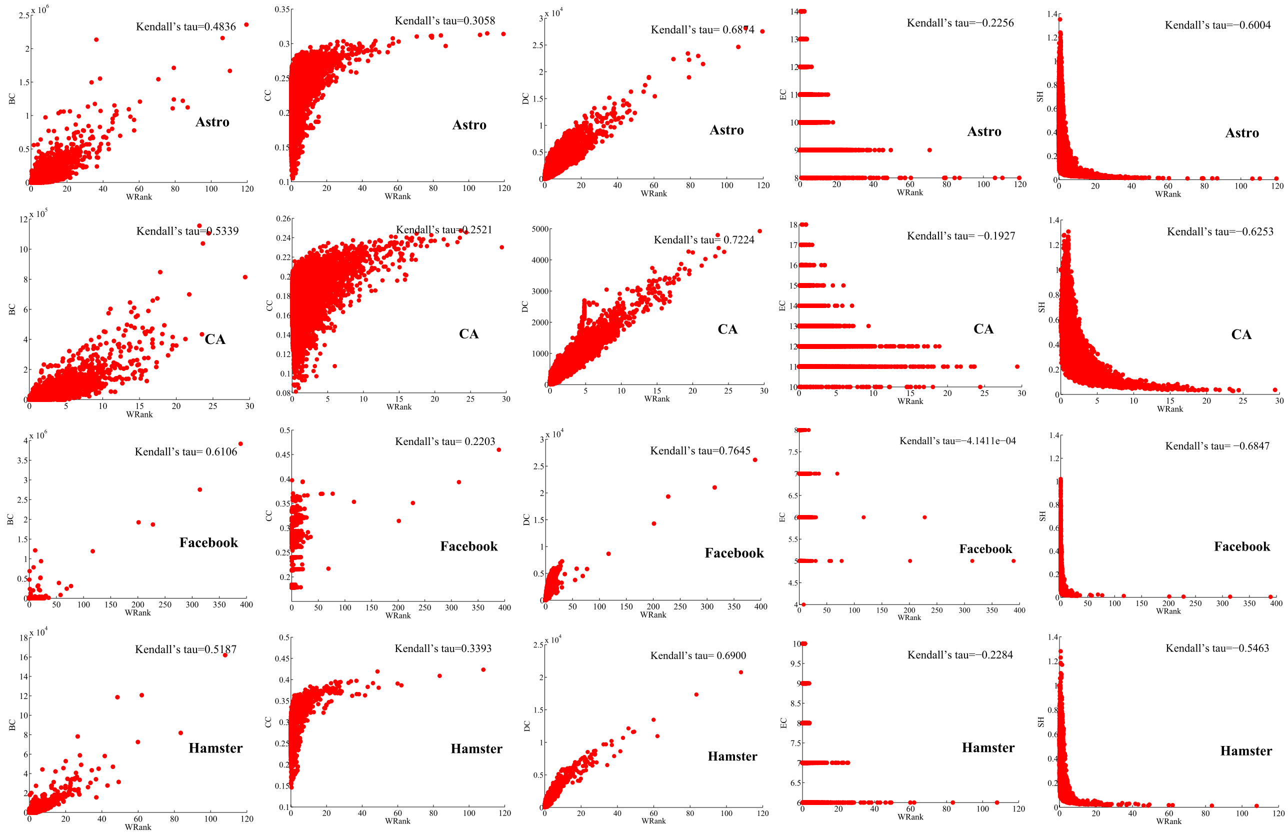

Figure 11. Correlation analysis experiments in weighted-undirected networks.

$$\tau = \frac{N_c - N_d}{N(N-1)/2} \tag{9}$$

where N_c and N_d are the number of concordant and discordant pairs, respectively. N is the number of nodes in the network.

In undirected networks, from Figs. 10 and 11, we can see that our centrality index is negatively correlated with EC and SH, because EC considers the node with the largest distance from the node, while SH considers the constraint coefficient of the node. The smaller the constraint coefficient is, the more important the node is, contrary to our centrality index in this paper. In unweighted-undirected networks, we can see from Fig. 10 that there is no obvious correlation between our centrality index and other centralities, but in the Facebook network, our centrality index has a high positive correlation with BC, CC and DC. The reason may be that Facebook is a social network, and the propagation between nodes is similar to that of the adjacency entropy algorithms in this paper. In weighted-undirected networks (Fig. 11) and directed networks (Supplementary Figs. S7 and S8), we can clearly observe a high correlation between our centrality index and DC. The reason may be that our centrality index and DC are designed based on the local properties of nodes. Similarly, we can find that our centrality index has low correlation with BC, CC and EC in the four different types of networks because BC, CC and EC are global metrics. By comparing the correlations between our centrality index and PR, VC and other centralities in directed networks (Supplementary Figs. S7 and S8), we can find that the correlations between our centrality index and PR and VC are greater than those of other centralities (except DC) because PR and VC are both centralities designed for directed networks, while other centralities are applicable to both directed networks and undirected networks.

Computational efficiency.

The adjacency information entropy algorithm has two steps: the calculation of the adjacency degree and the adjacency information entropy. Since every node's adjacency degree and adjacency information entropy in the network needs to be calculated, the computational complexity of the first cycle is $O(N)$, where N is the number of the network nodes. In the calculation of the node adjacency degree, it is also necessary to traverse the neighbouring nodes of the network nodes; thus, the total computational complexity of our algorithm is $O(N^2)$. Since the metric used in this paper involves the first-order neighbour of the node, the algorithmic complexity is lower than those of the global metrics, such as betweenness centrality(BC) with complexity $O(MN^3)$ and closeness centrality(CC) with complexity $O(MN^2)$, where M is the number of edges in the network. The number of network nodes applied by our algorithm could be further scaled up under the High Performance Computing (HPC) environment.

Conclusion

In this paper, we design two vital node identification algorithms for four different types of networks. By calculating and comparing the adjacency information entropy of nodes, the importance of nodes is ranked. The larger the entropy value is, the more vital the nodes are. The algorithms highlight the different characteristics of the different types of networks. For weighted networks, the strength of the nodes is used to calculate the adjacency information entropy instead of the degree of the nodes. For directed networks, the influence coefficient of a node's in-degree and out-degree value is used, which further refines the influence of a node's in-degree and out-degree on the node's importance. The experimental results show that our proposed algorithms outperform several benchmark methods. In the future, we will consider identifying vital nodes for more realistic network types, including temporal networks, etc.

Received: 30 September 2019; Accepted: 2 February 2020;
Published online: 14 February 2020

References

1. Kitsak, M. *et al*. Identification of influential spreaders in complex networks. *Nature Physics* **6**, 888–893 (2010).
2. Freeman, L. C. Centrality in social networks conceptual clarification. *Social Networks.* **1**, 215–239 (1978).
3. Hage, P. & Harary, F. Eccentricity and centrality in networks. *Social Networks.* **17**, 57–63 (1995).
4. Sabidussi, G. The Centrality Index of a Graph. *Psychometrika* **31**, 581–603 (1996).
5. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1997).
6. Shimbel, A. Structural parameters of communication networks. *Bulletin of Mathematical Biophysics* **15**, 501–507 (1953).
7. Shaw, M. E. Group Structure and the Behavior of Individuals in Small Groups. *Journal of Psychology Interdisciplinary & Applied* **38**, 139–149 (1954).
8. Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology* **2**, 113–120 (1972).
9. Brin, S. & Page, L. The anatomy of a large-scale hypertextual Web search engine. *International Conference on World Wide Web* **1**, 107–117 (1998).
10. Gino, D. F. *et al*. Finding influential nodes for integration in brain networks using optimal percolation theory. *Nature Communications* **9**, 2274–2286 (2018).
11. Chen, D., Lü, L., Shang, M. S., Zhang, Y. C. & Zhou, T. Identifying influential nodes in complex networks. *Physica A: Statistical Mechanics & its Applications* **391**, 1777–1787 (2012).
12. Liu, Y., Tang, M., Zhou, T. & Do, Y. Identify influential spreaders in complex networks, the role of neighborhood. *Physica A: Statistical Mechanics & its Applications* **452**, 289–298 (2015).
13. Yu, H., Cao, X., Liu, Z. & Li, Y. Identifying key nodes based on improved structural holes in complex networks. *Physica A: Statistical Mechanics & its Applications* **486**, 318–327 (2017).
14. Zhang, J. X., Chen, D. B., Dong, Q. & Zhao, Z. D. Identifying a set of influential spreaders in complex networks. *Scientific Reports* **6**, 27823–27833 (2016).
15. Ma, Q. & Ma, J. Identifying and ranking influential spreaders in complex networks with consideration of spreading probability. *Physica A: Statistical Mechanics & its Applications* **465**, 312–330 (2017).

16. Lü, L., Chen, D., Ren, X. L., Zhang, Q. M., Zhang, Y. C. & Zhou, T. Vital nodes identification in complex networks. *Phys. Rep.* **650**, 1–63 (2016).
17. Lü, L., Zhang, Y. C., Yeung, C. H. & Zhou, T. Leaders in Social Networks, the Delicious Case. *PLOS One* **6**, e21202 (2011).
18. Min, B. Identifying an influential spreader from a single seed in complex networks via a message-passing approach. *The European Physical Journal B* **91**, 18–24 (2018).
19. Liu, J. G., Lin, J. H., Guo, Q. & Zhou, T. Locating influential nodes via dynamics-sensitive centrality. *Scientific Reports* **6**, 21380–21388 (2016).
20. Zhang, J., Xu, X. K., Li, P., Zhang, K. & Small, M. Node importance for dynamical process on networks: A multiscale characterization. Chaos: An Interdisciplinary. *Journal of Nonlinear Science* **21**, 47–4 (2011).
21. Gao, C., Wei, D., Hu, Y., Mahadevan, S. & Deng, Y. A modified evidential methodology of identifying influential nodes in weighted networks. *Physica A: Statistical Mechanics & its Applications* **392**, 5490–5500 (2013).
22. Wei, D., Deng, X., Zhang, X., Deng, Y. & Mahadevan, S. Identifying influential nodes in weighted networks based on evidence theory. *Physica A: Statistical Mechanics & its Applications* **392**, 2564–2575 (2013).
23. Eidsaa, M. & Almaas, E. S-core network decomposition: A generation of k-core analisis to weighted networks. *Phys. Rev. E.* **88**, 062819 (2013).
24. Chen, D. B., Gao, H., Lü, L. & Zhou, T. Identifying Influential Nodes in Large-Scale Directed Networks: The Role of Clustering. *Plos One.* **8**, e77455 (2013).
25. Yang, Y., Xie, G. & Xie, J. Mining Important Nodes in Directed Weighted Complex Networks. *Discrete Dynamics in Nature and Society*. 1–7(2017).
26. Edgar, A. P. W., Sooyeon, Y., Antonio, L. F., Jose, F. F. M. & Alexander, V. G. The central role of peripheral nodes in directed network dynamics. *Scientific Reports* **9**, 2045–2322 (2019).
27. Wang, Y. & Liu, J. G. Evaluation method of node importance in directed-weighted complex network based on multiple influence matrix. *Acta Physica Sinica* **66**, 13–24 (2017).
28. Newman, M. E. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 404–409 (2001).
29. Leskovec, J., Kleinberg, J. & Faloutsos, C. Graph evolution: Densification and shrinking diameters. *Acm Transactions on Knowledge Discovery from Data* **1**, 2 (2007).
30. Kunegis, J. Hamsterster full network dataset-KONECT.Available at, http://konect.uni-koblenz.de/networks/petster-hamster (Accessed:01/03/2014).
31. Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F. & Arenas, A. Self-similar community structure in a network of human interactions. *Physical Review E* **68**, 065103 (2004).
32. Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A. & Arenas, A. Models of social networks based on social distance attachment. *Physical Review E Statistical Nonlinear & Soft Matter Physics* **70**, 056122 (2004).
33. Spring, N., Mahajan, R. & Wetherall, D. Measuring ISP topologies with Rocketfuel. *IEEE/ACM Trans. Netw.* **12**, 2 (2004).
34. Burt, R. S. Structural Holes: The Social Structure of Competition, *Harvard University Press* (2009).
35. Watts, D. J. & Strogatz, S. H. Collective dynamics of ’small-world". *Nature* **393**, 440–442 (1998).
36. Kendall, M. G. The treatment of ties in ranking problems. *Biometrika* **33**, 239–251 (1945).

Acknowledgements

The work is partially supported by National Natural Science Foundation of China (71571186, 61703416), Natural Science Foundation of Hunan Province, China (2018JJ3614), Postgraduate Research Innovation Project of Hunan Province, China (CX20190041)

Author contributions

X.X., C.Z. and Q.Y.W. devised the research project. X.X. and Q.Y.W. performed the research and analyzed the data. X.X., C.Z., Q.Y.W., Y.Z. and X.Q.Z. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-59616-w.

Correspondence and requests for materials should be addressed to X.X. or C.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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