Pramana – J. Phys. (2019) 92:68
https://doi.org/10.1007/s12043-019-1727-1

A modified efficiency centrality to identify influential nodes in weighted networks

YUNCHUAN WANG[1], SHASHA WANG[2] and YONG DENG[3,4,5],*

[1]School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]School of Computer Science, Sichuan University, Chengdu, China
[3]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China
[4]Big Data Decision Institute, Jinan University, Guangzhou, China
[5]School of Computer and Information Science, Southwest University, Chongqing, China
*Corresponding author. E-mail: dengentropy@uestc.edu.cn

MS received 23 February 2018; revised 22 August 2018; accepted 11 September 2018; published online 28 February 2019

Abstract. It is still a crucial issue to identify influential nodes effectively in the study of complex networks. As for the existing efficiency centrality (EffC), it cannot be applied to a weighted network. In this paper, a modified efficiency centrality ($\mathrm{EffC^m}$) is proposed by extending EffC into weighted networks. The proposed measure trades off the node degree and global structure in a weighted network. The influence of both the sum of the average degree of nodes in the whole network and the average distance of the network is taken into account. Numerical examples are used to illustrate the efficiency of the proposed method.

Keywords. Complex network; influential nodes; weighted network; efficiency centrality.

PACS No. 02.90.+p

1. Introduction

In recent years, the complex network has attracted much attention in numerous fields, such as economics, management, computer science, natural sciences, biological sciences etc. [1–12]. It is a critical task to find an effective method to identify influential nodes in complex networks because of their theoretical significance and practical value, including controlling rumours, spreading of diseases [13–15] and creating new marketing tools [1–3,16]. Besides, the centrality methods are also relevant to the text analysis and scientific science, because they are closely related to the definition of efficient and robust metrics [17–22].

Various centrality measures have been proposed over the past years to capture the rank of a complex network according to the degree and weight strength of nodes as well as topological importance in the network structure [23]. Three commonly used methods in analysing the nodes centrality are: degree centrality (DC), betweenness centrality (BC) [24] and closeness centrality (CC) [25]. The DC method is very simple and straightforward but of little relevance, because it only takes into account the degree of nodes but neglects the global structure of the network. The BC and CC are celebrated global metrics which can better identify the influential nodes that make a difference in networks. However, they cannot be applied in large-scale networks due to their computational complexity [26]. Another limitation of CC which is the lack of applicability to networks with disconnected components: two nodes that belong to different components but do not have a finite distance between them.

Except these three measures, several available centrality measures are also proposed, such as semilocal centrality [26], eigenvector centrality [27], Katz's centrality [28], PageRank [29] and LeaderRank [30]. These methods do have great performances in an unweighted network but do not work so well in a weighted network, because they are just designed for unweighted and undirected networks at the beginning [31]. However, there are lots of weighted networks in real world [32,33]. Some centrality measures have been extended to weighted networks [14,24,34,35]. In addition, due to the efficiency to combine different data [36–38], evidence theory is applied to take the different measures into consideration to obtain the final result [39–42]. It is still an open issue to design an effective ranking measurement to capture influential nodes.

The existing efficiency centrality (EffC) [43] based on network efficiency is obtained by considering the influence of each node's contributions to the entire network's efficiency. This method performs well in many unweighted networks. However, it cannot be applied in weighted networks because the definition of network efficiency was proposed for unweighted networks at the beginning, and it neglects the weight strength of nodes in the weighted networks. In this paper, we propose the network efficiency in a weighted network [44]. Then the EffC in a weighted network ($\mathrm{EffC}^{\mathrm{m}}$) is proposed by a combination of degree and weight strength of each node. To evaluate the performance of the proposed centrality measure, we adopt the susceptible and infected (SI) model to examine the spreading influence of the nodes ranked by different centrality measures. The simulations on real networks are used to show the efficiency of the proposed method.

The paper is organised as follows. Section 2 begins with a brief overview of the existing centrality measures. Then, the proposed method for identifying the influential nodes is developed and illustrated by an example network in §3. In §4, the SI model is used to evaluate the performance in a real complex network. Finally, a simple conclusion is presented in §5.

2. Preliminary

2.1 *Centrality measures for influential nodes*

An unweighted network can generally be represented as a set $G = (V, E)$. Here, V and E represent the number of nodes and the number of edges, respectively. As for a weighted network, it is described as a set $G = (V, E, W)$ [25]. W is the weight set of E, i.e., link E_{ij} from nodes i to j has a weight $\omega_{ij} \in W$.

DEFINITION 1

The DC of node i, denoted as d_i, is defined as

$$k_i = \sum_{j}^{N} x_{ij}, \tag{1}$$

where N is the total number of nodes i and x_{ij} represents the connection between nodes i and j, i is the focal node, j represents all other nodes. The value of x_{ij} is defined as 1 if node i is connected to node j, and 0 otherwise.

DEFINITION 2

The BC of node i, denoted as b_i, is given as

$$b_i = \sum_{j,k \neq i} \frac{g_{jk}(i)}{g_{jk}}, \tag{2}$$

where g_{jk} denotes the number of shortest binary paths between nodes j and k, and $g_{jk}(i)$ is the number of those paths that go through node i [25].

DEFINITION 3

The CC of node i, denoted as c_i, is defined as

$$c_i = \frac{1}{\sum_j^N d_{ij}}, \tag{3}$$

where d_{ij} denotes the shortest distance from node i to node j.

The sum of the weights between two nodes represents the distance between them in a weighted network. Here are the above measures which are extended to weighted networks [24].

DEFINITION 4 (DC in a weighted network)

The DC of node i, denoted as d_i^{w}, is defined as

$$k_i^{\mathrm{w}} = \sum_{j}^{N} \omega_{ij}, \tag{4}$$

where ω_{ij} is the weight of the edges between nodes i and j, which is greater than 0 when node i is connected to node j.

DEFINITION 5 (BC in a weighted network)

The BC [24] of node i, denoted as b_i^{w}, is defined as

$$b_i^{\mathrm{w}} = \sum_{j,k \neq i} \frac{g_{jk}^{\mathrm{w}}(i)}{g_{jk}^{\mathrm{w}}}, \tag{5}$$

where g_{jk}^{w} represents the number of binary shortest paths between nodes j and k, and $g_{jk}^{\mathrm{w}}(i)$ is the number of those paths that go through node i.

DEFINITION 6 (CC in a weighted network)

The CC [25] of node i, denoted as $c^{\mathrm{w}}(i)$, is defined as

$$c_i^{\mathrm{w}} = \frac{1}{\sum_j^N d_{ij}}, \tag{6}$$

where d_{ij} is the shortest distance or connection strength between nodes i and j.

2.2 *Efficiency centrality*

EffC [43] takes into account both the influence of each node itself and the global structure of the graph. The network efficiency is a measurement of how efficiently the information passes within the nodes. The kernel of EffC is that removing a single node in an unweighted network, then assessing the efficiency changes of the network. Once a node is removed in a network, at the same time, the edges that are related to it will disappear. Therefore, if a pivotal node is removed, the structure and efficiency of the network will change a lot (e.g. the changing of the shortest path between two nodes, the connectivity of a graph).

DEFINITION 7

The efficiency e_{ij} is defined as [43]

$$e_{ij} = \frac{1}{d_{ij}}, \tag{7}$$

where d_{ij} is the shortest distance between nodes i and j.

DEFINITION 8

Denote $E[G]$ as the efficiency of network G [43]

$$E[G] = \frac{\sum_{i \neq j \in G} e_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}, \tag{8}$$

where $E[G]$ is the average of e_{ij} and measures the mean flow rate of information over G and the quantity of $E[G]$ varies in the range [0,1].

DEFINITION 9

The EffC of node k is defined as

$$\text{EffC}(k) = \frac{\Delta E}{E} = \frac{E[G] - E[G'(k)]}{E[G]}, \quad k = 1, \ldots, N. \tag{9}$$

Here, if the node k has Q links to neighbours and the sum of edges is P, the subgraph $G'(k)$ indicates a graph with $N-1$ nodes and $P-Q$ edges obtained by removing node k and its neighbour edges from G.

3. Modified efficiency centrality

In unweighted networks, the efficiency of a network is defined as the average of the shortest distance of the network as shown in eq. (8). It gains good results in unweighted network. However, if we extend this definition to weighted networks, the results seem to be inaccurate because it considers the degree of nodes only. In weighted complex networks, the distance between two nodes is represented by the sum of the weights between the nodes. In this paper, we modify the definition of efficiency in weighted networks. Both the degree and the weight of a node are considered, and the global structure is concerned about at the same time. The modified EffC conforms to the ideas of the EffC in unweighted networks [43]. Each node will be removed, and the affecting degree of network efficiency and structure will be calculated after each removal. Obviously, once a node disappears in a network, the edges related to the node will disappear. The removal of a node in a weighted network will have a different impact depending on the node itself. If an important node in the network is removed, its removal will bring great change to the network efficiency and global structure. Inspired by network aggregation degree [44], we extend EffC to be applied in weighted networks.

DEFINITION 10

Denoting s as the sum of the average degree of nodes in a weighted network

$$s = \sum_{1}^{N} \frac{1}{N_i} \cdot \sum_{j \in N_i} \frac{1}{\omega_{ij}}, \tag{10}$$

where N_i is the set of the neighbours of node i. It takes into account the nodes and their neighbours and the weight between them.

DEFINITION 11

Denoting L^{m} as the modified average distance of a network [44],

$$L^{\mathrm{m}} = \frac{1}{N(N-1)} \cdot \sum_{i \neq j} d_{ij}^{\mathrm{m}}. \tag{11}$$

The number of nodes in the network is denoted by N and d_{ij}^{m} denotes the average distance between nodes i and j in the degraded network which is an unweighted network keeping the routes of the weighted network. By this modification, the distinct inaccuracy that a node which has less number of neighbours than others will show greater importance, can be avoided.

DEFINITION 12

Denoting $E^{\mathrm{m}}[G]$ as the efficiency of the weighted network G,

$$E^{\mathrm{m}}[G] = \frac{1}{s \cdot L^{\mathrm{m}}} = \frac{1}{\sum_1^N \frac{1}{N_i} \cdot \sum_{j\in N_i} \frac{1}{\omega_{ij}} \cdot \frac{1}{N(N-1)} \cdot \sum_{i\neq j} d_{ij}^{\mathrm{m}}}. \tag{12}$$

When $N = 1$, let $E^{\mathrm{m}}[\mathrm{G}] = 1$, and so the efficiency of the weighted network G varies in the range [0,1].

DEFINITION 13

The modified EffC in a weighted network, $\mathrm{EffC}^{\mathrm{m}}$, is defined as

$$\mathrm{EffC}^{\mathrm{m}}(k) = \frac{\Delta E^{\mathrm{m}}}{E^{\mathrm{m}}} = \frac{E^{\mathrm{m}}[G] - E^{\mathrm{m}}[G'(k)]}{E^{\mathrm{m}}[G]}, \quad k = 1, \ldots, N. \tag{13}$$

Here, if the node k has Q links to neighbours and the sum of edges is E, $G'(k)$ means a graph with $N - 1$ nodes and $E - Q$ edges obtained by removing node k and its neighbour edges from G.

3.1 *Explanation of the example*

The network analysis is widely used to model complex systems [45–47]. In this section, a numerical example is given to illustrate the proposed method $\mathrm{EffC}^{\mathrm{m}}$. We consider a network with 8 nodes and 13 weighted edges which are shown in figure 1, where the thickness of the edge indicates the size of the weight and the weight of the unlabelled edges is 1. From eqs (10) to (13), the initial attributes of the example network are calculated as follows:

$$s = \sum_1^8 \frac{1}{N_i} \cdot \sum_{j\in N_i} \frac{1}{\omega_{ij}} = 7.0119,$$

$$L^{\mathrm{m}} = \frac{1}{8(8-1)} \cdot \sum_{i\neq j} d_{ij}^{\mathrm{m}} = 1.5536,$$

$$E^{\mathrm{w}}[G] = \frac{1}{s \cdot L^{\mathrm{m}}} = \frac{1}{\sum_1^8 \frac{1}{N_i} \cdot \sum_{j\in N_i} \frac{1}{\omega_{ij}} \cdot \frac{1}{8(8-1)} \cdot \sum_{i\neq j} d_{ij}^{\mathrm{m}}} = 0.0918. \tag{14}$$

When node 4 is removed, the subgraph is as shown in figure 2, where the dashes represent the removed part. Then calculate each value as follows:

Firstly, calculate the sum of the average degree and the modified average distance of the whole network after removing node 4, respectively, with eqs (10) and (11):

$$s = \sum_1^7 \frac{1}{N_i} \cdot \sum_{j\in N_i} \frac{1}{\omega_{ij}} = 6.000,$$

$$L^{\mathrm{m}} = \frac{1}{7(7-1)} \cdot \sum_{i\neq j} d_{ij}^{\mathrm{m}} = 1.8333.$$

Secondly, obtain the efficiency of the network after the removal by eq. (12):

$$E^{\mathrm{w}}[G'(4)] = \frac{1}{s \cdot L^{\mathrm{m}}} = \frac{1}{\sum_1^7 \frac{1}{N_i} \cdot \sum_{j\in N_i} \frac{1}{\omega_{ij}} \cdot \frac{1}{7(7-1)} \cdot \sum_{i\neq j} d_{ij}^{\mathrm{m}}} = 0.0909.$$

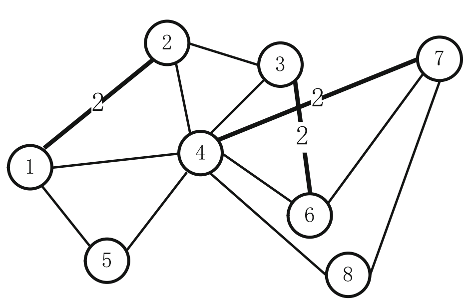

Figure 1. A weighted network with 8 nodes and 13 edges (the weight of the unlabelled edges is 1).

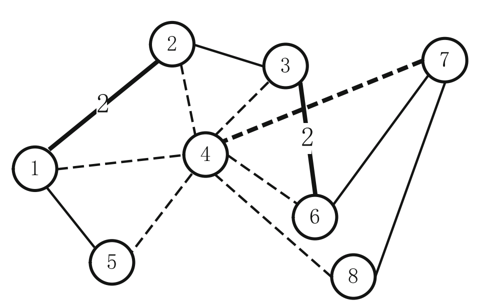

Figure 2. A weighted network after removing node 4.

Table 1. $\mathrm{EffC^m}$ of each node in figure 1. The first row shows the initial attributes of the whole network without removing any node. s is the sum of the average degree of nodes of graph 1. L^{m} is the modified average distance of the network.

Removed node i	s	L^{m}	$E^{\mathrm{m}}[G]$	$\mathrm{EffC^m}(k)$
–	7.0119	1.5536	0.0918	–
Node 1	6.3333	1.5476	0.1020	−0.1114
Node 2	6.2500	1.5476	0.1034	−0.1262
Node 3	6.2500	1.5476	0.1034	−0.1262
Node 4	6.0000	1.8333	0.0909	0.0097
Node 5	5.9167	1.5000	0.1127	−0.2274
Node 6	6.3333	1.5238	0.1036	−0.1288
Node 7	6.2500	1.5238	0.1050	−0.1438
Node 8	5.7500	1.5000	0.1159	−0.2630

Finally, $\mathrm{EffC^m}$ of node 4 is calculated by eq. (13):

$$\mathrm{EffC^w}(4) = \frac{\Delta E^{\mathrm{w}}}{E^{\mathrm{w}}} = \frac{E^{\mathrm{w}}[G] - E^{\mathrm{w}}[G'(4)]}{E^{\mathrm{w}}[G]} = 0.0097.$$

The $\mathrm{EffC^m}$ values of the other nodes are similarly obtained and they are shown in the fifth column of table 1.

As can be seen in table 1, the value of $\mathrm{EffC^m}$ of node 4 is the maximum of all nodes in the example network. In fact, it is intuitional that node 4 is a pivotal node in the network because it connects to most nodes. Without node 4, the network efficiency will decline obviously. To better analyse the effect of this proposed method on identifying influential nodes in networks, it is applied in the real complex networks.

4. Applications and analysis

4.1 *The EffC^m applied to real networks*

In this section, some real complex networks are used to evaluate the performance of the proposed measure. The following is the detailed description of the chosen networks:

(i) Freeman's electronic information exchange system (EIES) network data [24,25]

The dataset was collected from three different network relations among researchers by Freeman in 1978. It is a weighted and directed network which contains three networks of researchers who is working on social network analysis. We adopt the third network. The nodes in the third network denote the number of messages sent among 32 researchers on an electronic communication tool, and the node's strength in this network is based on a ratio scale [24]. For instance, the degree of node Lin Freeman is 31 and the weight is 3171. It means that he had contact with 31 other researchers, and the total number of messages sent in electronic communication tools is 3171.

(ii) USAir97 network data

This is an undirected and weighted network, which is a US air flight network in 1997. In the network, each node represents a US airport, and each line contains the name of two US airports. The weights between two nodes denote the distances between two connected airports in the coordinate axis. It can be downloaded on 'http://vlado.fmf.uni-lj.si/pub/networks/data/'.

(iii) Groad network data

This is a network describing the link information of German highway network. It includes the adjacency matrix and the labels of all 1168 nodes [48]. The data can be obtained on 'http://vlado.fmf.uni-lj.si/pub/networks/data/'.

(iv) Newman's scientific collaboration network

It is the co-authorship network based on preprints posted to Condensed Matter section of arXiv E-Print Archive between 1995 and 1999 [49]. It contains four networks of authors and papers. The third network is adopted, which is a weighted static one-mode network. The nodes in the third network denote the authors and the strength of the node in the network represents the sum of joint papers between different authors. The data can be obtained on 'https://toreopsahl.com/datasets/#newman2001/'. Table 2 shows the basic topological properties of the four networks.

4.2 *SI model for evaluation*

To evaluate the performance of our ranking method, the SI model [52,53] is adopted to examine the spreading

Table 2. The basic topological features of the four real networks. n and m are the total number of nodes and links, respectively. $k_{\max}$ and $\langle k\rangle$ denote the maximum and average degree. C and r are the clustering coefficient [50] and assortative coefficient [51], respectively.

Network	n	m	$\langle k\rangle$	$k_{\max}$	C	r
Freeman's	32	230	14.375	31	0.7332	−0.0320
USAir97	332	2126	12.810	139	0.3126	−0.2079
Groad	1168	1243	1.060	12	0.0012	0.0251
Newman's	16726	47594	5.691	107	0.0004	−0.1760

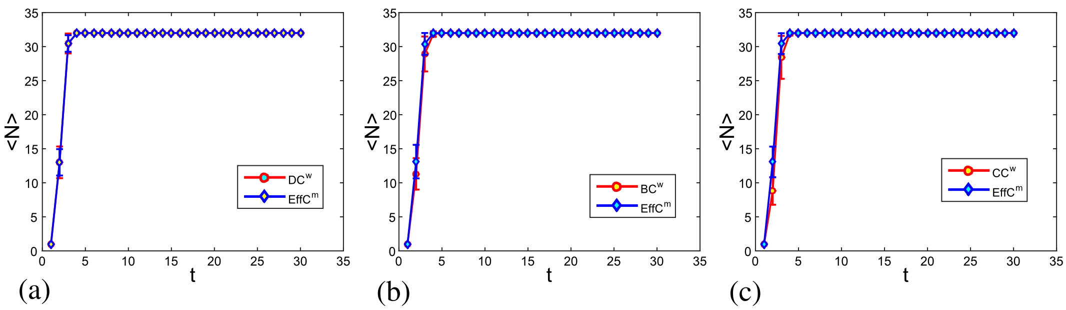

Figure 3. In Freeman's network, the cumulative number of infected nodes as a function of time with 30 steps. Results are obtained by averaging over 1000 implementations when $\alpha = 0.2$.

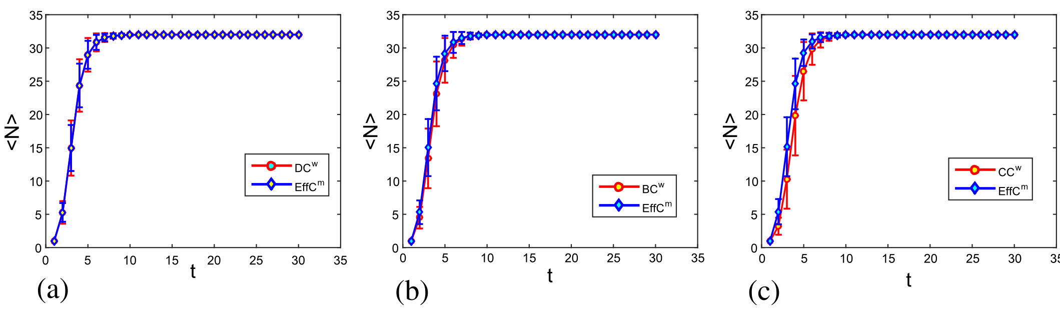

Figure 4. In Freeman's network, the cumulative number of infected nodes as a function of time with 30 steps. Results are obtained by averaging over 1000 implementations when $\alpha = 0.5$.

influence of top-L nodes ranked by these different measures.

In the SI model, there are two compartments, namely, susceptible $S(t)$ and infected $I(t)$. $S(t)$ is used to represent the number of individuals susceptible to the disease (not yet infected), while $I(t)$ means the number of individuals who have been confirmed to be infected and are capable of spreading the disease to other susceptible individuals. At the beginning, a node that is tested is set to be an infected one, and the infected individual will randomly infect susceptible neighbours with probability P at each step. Node j is infected by node i with probability P [54–56]:

$$P = \left[\frac{\omega_{ij}}{\omega_M + 1}\right]^{\alpha}, \quad \omega > 0, \tag{15}$$

where ω_{ij} is the weight of edge E_{ij} and α is a positive constant in the network. The smaller the α is, the more quickly the infection spreads because $[\omega_{ij}/(\omega_M + 1)] < 1$. How to evaluate the efficiency of identifying influential nodes is an open issue [57]. This model is slightly different from the standard SI model where all the neighbours of an infected node have the chance to be infected and it is used to mimic the limited spreading capability of an individual [26,58,59]. The trial proceeds until all nodes are infected. Denoting N_t as the sum of total infected nodes at time t, clearly N_t increases with t and remains stable when there is no node to be infected. A total of 1000 implementations are considered, and the mean value of N_t is denoted $\langle N_t \rangle$.

4.3 *The value of* α

In this section, we discuss the impact of the value of α on the assessment based on the SI model. According to eq. (15), the propagation probability P is determined by the parameter α. The change of the parameter α will cause the changes of the infection efficiency of infected nodes at each time. As $[\omega_{ij}/(\omega_M + 1)] < 1$, the smaller the α is, the more quickly the infection spreads.

We focus on whether the values of α will have an effect on the comparison of the two methods. The Freeman's EIES network data is adopted to evaluate the effect of the value of parameter α. As shown in figures 3–6, the change of parameter α will mainly affect the propagation time (steps) when all the nodes in the network are infected. The smaller the α is, the more quickly the infection spreads, i.e., the shorter the time needed for all nodes to be infected. Such changes are global and it will affect the infection efficiency of all infected nodes. The difference in the infection capacity of nodes still depends on their own information (degree and weights) and their location in the network. If the value of α is too small, it will not result in any significant difference in

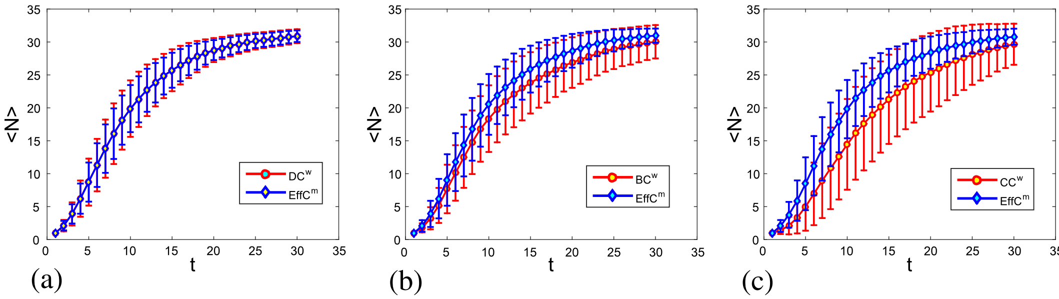

Figure 5. In Freeman's network, the cumulative number of infected nodes as a function of time with 30 steps. Results are obtained by averaging over 1000 implementations when $\alpha = 1$.

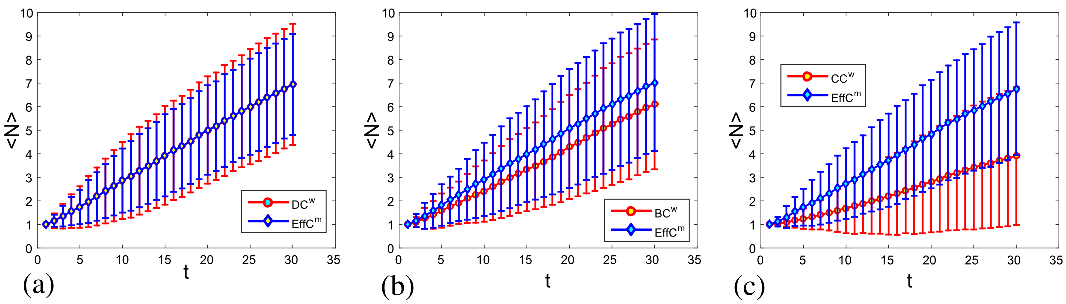

Figure 6. In Freeman's network, the cumulative number of infected nodes as a function of time with 30 steps. Results are obtained by averaging over 1000 implementations when $\alpha = 2$.

Table 3. The rank of nodes obtained by different measures.

Rank	DC	BC	CC	EffC^{m}	DC	BC	CC	EffC^{m}
	Freeman				*USAir97*			
1	1	8	5	1	118	8	172	118
2	29	11	24	29	261	118	182	47
3	8	24	8	31	248	261	198	65
4	2	32	30	2	67	47	216	144
5	31	31	32	8	255	313	166	33
6	32	29	4	11	47	201	206	147
7	11	2	6	24	166	13	225	162
8	24	18	18	32	201	67	133	311
9	10	25	17	27	182	182	140	219
10	27	10	9	10	147	255	181	177
	Groad				*Newman*			
1	693	219	698	543	755	36	–	755
2	403	543	219	393	1846	1267	–	4474
3	300	698	450	850	80	213	–	1846
4	217	693	565	853	1842	1295	–	7315
5	373	758	331	219	1530	755	–	4034
6	410	403	763	861	1529	4474	–	5489
7	758	763	267	531	208	4034	–	311
8	207	565	663	889	311	2084	–	80
9	219	886	729	909	1714	52	–	1842
10	331	373	347	511	1713	406	–	7314

the propagation effect. With the increase of α, the difference in infection efficiency is more significant. But the value of α may be too large to lead to more infection within the specified time. Therefore, the value of α does not affect the final comparison of different measurement methods. It can be seen that under the same α, the top-10 nodes evaluated by effective centrality can always infect all nodes faster and effectively.

4.4 *Experimental results and analysis*

In this section, different real networks are used to illustrate the difference between the proposed measures EffC^{m} and DC, BC, CC. The DC, BC and CC methods are extended to be applied in weighted networks, denoted as DC^{w}, BC^{w} and CC^{w}, and given by eqs (4)–(6). The top-10 list obtained by different centrality measures is shown in table 3. As shown in table 3, both DC and EffC^{m} rank node 1 as the top-1 node in Freeman's EIES network, while BC identifies node 8 as the influential spreaders. However, CC ranks node 5 as the top-1 node. In USAir97 network, node 118 is selected as the most influential node by DC and EffC^{m}, and BC ranks it as the top-2 node. CC assigns the highest score to node 172. As the size of the network grows, the list of top-10 nodes found by different methods is quite different. In the Groad network, nodes 543 and 698 seem to perform better than other nodes by most centrality measures. Because the Newman network is directed, CC does not apply to this network. Thus, the column of CC in Newman is empty. DC and EffC^{m} assign the highest score to node 755. However, there is still a big difference between the elements and the order of the top-10 nodes evaluated by different methods. In order to evaluate the difference between our ranking method and others, we calculate the cumulative number of infected nodes with the initially infected nodes being those that appears in the top-10 list by the different centrality measures. The results are shown in figures 7–10.

From figures 7a–7c, one can observe that EffC^{m} is seen to be the best among all four measures in Freeman's network. EffC^{m} performs better than BC^{w} and CC^{w}. In figure 7a, it seems that the EffC^{m} measure behaves quite similar or close to DC. However, it can be seen that the number of errors in EffC^{m} is smallest compared to DC^{w}, BC^{w} and CC^{w}. That is, the calculation of EffC^{m} is more accurate.

In USAir97 network (figure 8), it can be seen that the final number of our proposed EffC^{m} is larger than the other three measures. In figures 8a and 8b, one can observe that the effect of the proposed method is always better than DC^{w} and BC^{w}. In figure 8c, although the node spreading speed after $t = 30$ is almost the same with CC^{w}, the propagation speed in the early stage is better than that of CC^{w}. In addition, it is shown in figure 8c that errors of EffC^{m} are smallest and so we can infer that the final number of the infected nodes is more effective and accurate. This inference indicates exactly that our proposed EffC^{m} is better.

In Groad network, as shown in figure 9, it can be seen that the contrast with BC and DC (figures 9a and 9b) shows that more nodes are infected after 80 steps, which are caused by EffC^{m}'s top-10 nodes. In figure 9c, the process of infection is almost the same with CC^{w} when $t > 50$, and the errors of EffC^{m} are obviously smaller than that of DC^{w} and BC^{w}.

From figure 10, one can see that EffC^{m} can achieve almost the same performance of infected nodes $F(t)$ compared to DC when $t < 10$ in the Newman's network, and EffC^{m} outperforms DC for the spreading ability while t is between 10 and 40. DC is slightly better than EffC^{m} after $t = 50$. EffC^{m} has better infectious effect and less error compared to BC as can be seen in figure 10b.

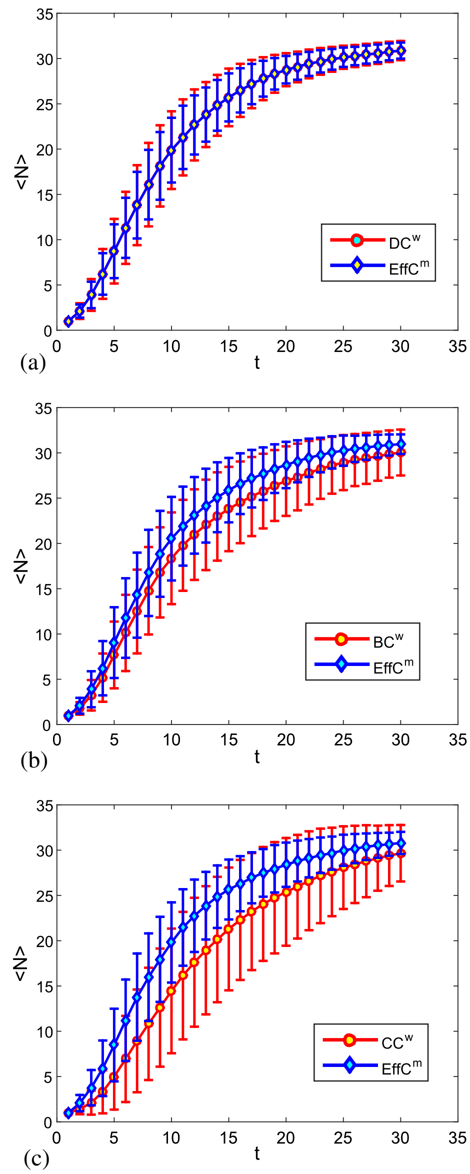

Figure 7. In Freeman's network, the cumulative number of infected nodes as a function of time with 30 steps, with the initially infected nodes being those that appear in the top-10 list by the proposed EffC^{m} and other centrality measures in Freeman's network. Results are obtained by averaging over 1000 implementations when $\alpha = 1$.

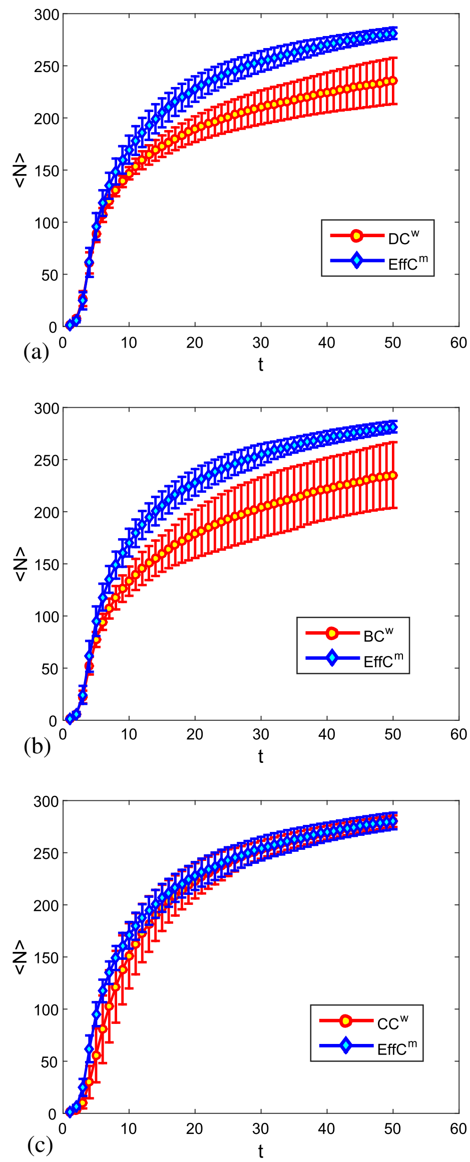

Figure 8. In USAir97 network, the cumulative number of infected nodes as a function of time with 50 steps, with the initially infected nodes being those that appear in the top-10 list by the proposed EffC$^{\mathrm{m}}$ and other centrality measures in Freeman's network. Results are obtained by averaging over 1000 implementations when $\alpha = 1$.

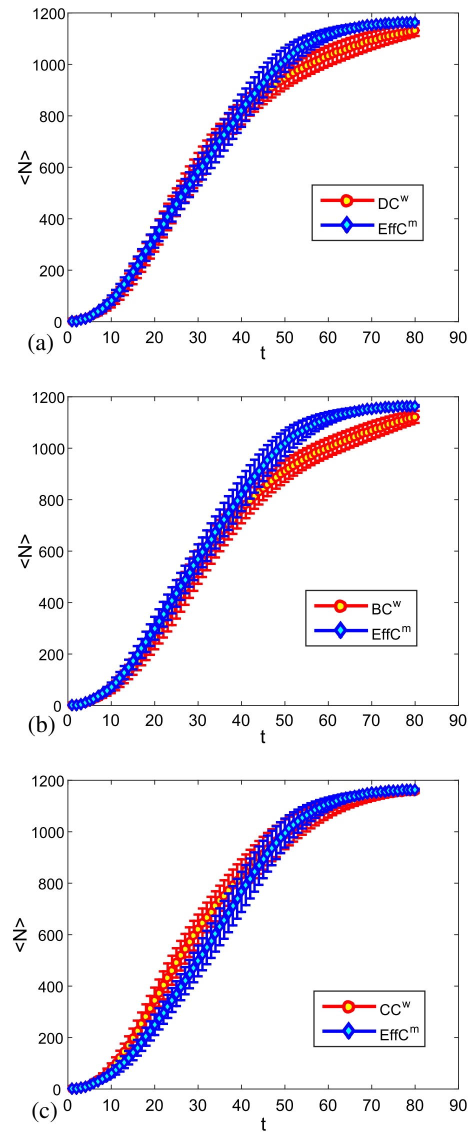

Figure 9. In the Groad network, the cumulative number of infected nodes as a function of time with 80 steps, with the initially infected nodes being those that appear in the top-10 list by the proposed EffC$^{\mathrm{m}}$ and other centrality measures in Freeman's network. Results are obtained by averaging over 1000 implementations when $\alpha = 1$.

4.5 *The EffC^{m} compared with other centrality measures*

In this section, four real network data are used to show the difference between the proposed method and DC$^{\mathrm{w}}$, BC$^{\mathrm{w}}$ and CC$^{\mathrm{w}}$. We mainly compare the influence of the nodes that appear in the top-10 list either by EffC$^{\mathrm{m}}$ or other centrality measures. From figures 7–9 it can be seen that the initially infected nodes which appear in the top-10 list by the proposed EffC$^{\mathrm{m}}$ can effectively infect other nodes compared to other centralities. That is, EffC$^{\mathrm{m}}$ can give comparatively better performance than others in identifying influential nodes.

Compared to other methods, EffC$^{\mathrm{m}}$ takes more information into account to capture the influential nodes. As can be seen in eqs (10)–(13), the definitions of s and L^{m} are based on DC, BC and CC. Both the degree of the node and the shortest paths between nodes are considered when ranking the nodes and the definition of E^{m} inherits the thought of EffC, removing a node and edges related to it to calculate the difference. In this way, the global structural information in a network is taken into consideration as well. One can observe

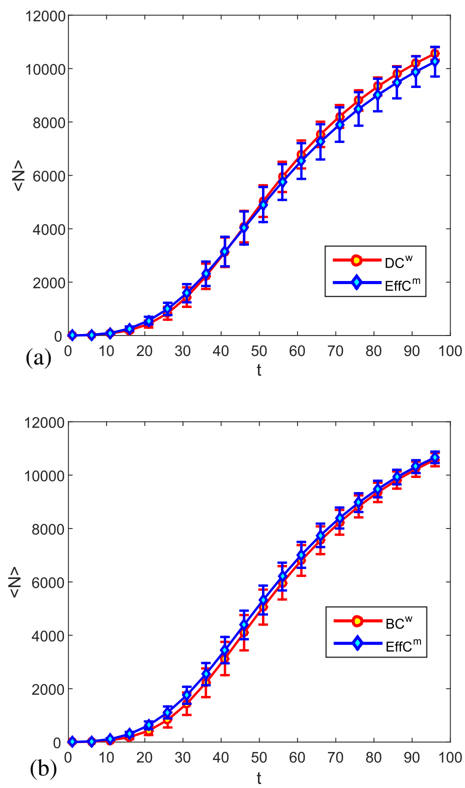

Figure 10. In the Newman network, the cumulative number of infected nodes as a function of time with 100 steps, with the initially infected nodes being those that appear in the top-10 list by the proposed EffC$^{\mathrm{m}}$ and other centrality measures in Freeman's network. Results are obtained by averaging over 1000 implementations when $\alpha = 1$.

that EffC$^{\mathrm{m}}$ is an improved method based on traditional measurements (DC, BC and CC) and EffC.

As the proposed EffC$^{\mathrm{m}}$ comprehensively considers the comprehensive information of nodes and networks based on DC, CC and EffC, the time complexity of the algorithm will increase accordingly. This method needs to calculate the shortest path of the network, which is calculated a total of N times. The calculation of CC takes computational complexity $O(n^4)$ with Floyd's algorithm [60]. It is a deficiency of this method. Although the computational complexity has increased, it can be seen from the evaluation results of the SI model that EffC$^{\mathrm{m}}$ can identify the influential nodes more efficiently and accurately. Therefore, EffC$^{\mathrm{m}}$ can be effectively adapted to capture influential nodes in small and medium networks, such as a traffic network.

5. Conclusion

In this paper, EffC$^{\mathrm{m}}$ measure is proposed based on weighted network efficiency. We define the modified efficiency of a network, and the comprehensive information of the degree of the node, the shortest distance between the nodes and the topology of the nodes in the network are taken into account in the proposed measure. To evaluate the performance, we apply our measure on four real networks and use SI model to evaluate the spreading process. Further, the average number of $F(t)$ $(t = 10)$ of the top-K nodes which are ranked by the proposed EffC and different centrality is used to compare and analyse the ranking ability of different measures.

Four different real networks are adopted to evaluate the effectiveness of the new approach. Through comparison of experiments with DC$^{\mathrm{w}}$, BC$^{\mathrm{w}}$ and CC$^{\mathrm{w}}$, our proposed EffC$^{\mathrm{m}}$ can give a comparatively better performance than the others. Node ranking results are based on more valid information, and so one can observe that EffC$^{\mathrm{m}}$ is more effective and accurate than the other methods from the comparison results of the SI model.

The numerical examples show that the proposed centrality can capture influential nodes well in networks as it considers more information than other methods. But the complexity of computing is one of its drawbacks. It is more suitable for identifying influential nodes of small- and medium-sized networks.

Acknowledgements

The authors greatly appreciate the reviewers' suggestions and the editor's encouragement. The work was partially supported by the National Natural Science Foundation of China (Grant Nos 61573290, 61503237).

References

[1] S H Strogatz, *Nature* **410**, 268 (2001)
[2] M E J Newman, *SIAM Rev.* **45(2)**, 167 (2003)
[3] E E Schadt, *Nature* **461(7261)**, 218 (2009)
[4] Y Yang, G Xie and J Xie, *Dis. Dyn. Nat. Soc.* **2017**, 1 (2017)
[5] S Wang, *Pramana – J. Phys.* **90(2)**: 25 (2018)
[6] F Nian and W Liu, *Pramana – J. Phys.* **86(6)**, 1209 (2016)
[7] Y C Lai, A Motter, T Nishikawa, K Park and L Zhao, *Pramana – J. Phys.* **64(4)**, 483 (2005)
[8] H E Ping, C G Jing, C Z Chen, T Fan and H S Nik, *Pramana – J. Phys.* **82(3)**, 499 (2014)
[9] D R Amancio, *Europhys. Lett.* **114(5)**, 58005 (2016)
[10] D R Amancio, *PLoS ONE* **10(8)**, e0136076 (2015)
[11] Y Han and Y Deng, *J. Ambient Intell. Humaniz. Comput.* **9(6)**, 1933 (2018)
[12] L Fei, Q Zhang and Y Deng, *Physica A* **512**, 1044 (2018)

[13] T Zhou, G Yan and B-H Wang, *Phys. Rev. E* **71(4)**, 046141 (2005)
[14] M Kitsak, L K Gallos, S Havlin, F Liljeros, L Muchnik, H E Stanley and H A Makse, *Nat. Phys.* **6(11)**, 888 (2010)
[15] D R Amancio, M G V Nunes, O N Oliveira Jr., T A S Pardo, L Antiqueira and L Da F Costa, *Phys. A: Stat. Mech. Appl.* **390(1)**, 131 (2011)
[16] Z Cheng, H T Zhang, M Z Q Chen, T Zhou and N V Valeyev, *PLoS ONE* **6(7)**, e22123 (2011)
[17] D R Amancio, *PLoS ONE* **10(2)**, e0118394 (2015)
[18] D R Amancio, *Scientometrics* **105(3)**, 1763 (2015)
[19] D R Amancio, O N Oliveira Jr and L Da F Costa, *New J. Phys.* **14(21)**, 1759 (2012)
[20] F N Silva, D R Amancio, M Bardosova, L Da F Costa and O N Oliveira Jr, *J. Informetr.* **10(2)**, 487 (2016)
[21] A Zeng, Z Shen, J Zhou, J Wu, Y Fan, Y Wang and H E Stanley, *Phys. Rep.* **714–715**, 1 (2017)
[22] M P Viana, D R Amancio and L Da F Costa, *J. Informetr.* **7(2)**, 371 (2013)
[23] V Nicosia, R Criado, M Romance, G Russo and V Latora, *Sci. Rep.* **2**, 218 (2012)
[24] T Opsahl, F Agneessens and J Skvoretz, *Soc. Netw.* **32(3)**, 245 (2010)
[25] L C Freeman, *Soc. Netw.* **1(3)**, 215 (1978)
[26] D Chen, L Lü, M S Shang, Y C Zhang and T Zhou, *Phys. A: Stat. Mech. Appl.* **391(4)**, 1777 (2012)
[27] P Bonacich and P Lloyd, *Soc. Netw.* **23(3)**, 191 (2001)
[28] L Katz, *Psychometrika* **18(1)**, 39 (1953)
[29] S Brin and L Page, *Comput. Netw. ISDN Syst.* **30**, 107 (1998)
[30] L Lü, Y C Zhang, H Y Chi and Z Tao, *PLoS ONE* **6(6)**, e21202 (2011)
[31] T Bian and Y Deng, *Chaos* **28**, 043109 (2018), https://doi.org/10.1063/1.5030894
[32] G Bagler, *Physica A* **387(12)**, 2972 (2008)
[33] R Zhang, B Ashuri and Y Deng, *Adv. Data Anal. Classif.* **11(4)**, 759 (2017)
[34] L Yin and Y Deng, *Physica A* **498**, 102 (2018)
[35] L Yin and Y Deng, *Physica A* **508**, 176 (2018)
[36] Y Li and Y Deng, *Int. J. Comput. Commun. Control* **13(5)**, 771 (2018)
[37] Y Han and Y Deng, *Soft Comput.* **22(15)**, 5073 (2018)
[38] H Mo and Y Deng, *Int. J. Fuzzy Syst.* **20(8)**, 2458 (2018), https://doi.org/10.1007/s40815-018-0514-3
[39] H Xu and Y Deng, *IEEE Access* **6(1)**, 11634 (2018)
[40] W Deng, X Lu and Y Deng, *Math. Probl. Eng.* **2018**, 1 (2018)
[41] T Bian, H Zheng, L Yin and Y Deng, *Qual. Reliab. Eng. Int.* **34**, 501 (2018)
[42] H Zheng and Y Deng, *Int. J. Intell. Syst.* **33(7)**, 1343 (2018)
[43] S Wang, Y Du and Y Deng, *Commun. Nonlinear Sci. Numer. Simul.* **47**, 151 (2017)
[44] T Zhu, S P Zhang, R X Guo and G C Chang, *Syst. Eng. Electron.* **31(8)**, 1902 (2009)
[45] M Li, Q Zhang and Y Deng, *Chaos Solitons Fractals* **117**, 283 (2018), https://doi.org/10.1016/j.chaos.2018.04.033
[46] B Kang, G Chhipi-Shrestha, Y Deng, K Hewage and R Sadiq, *Appl. Math. Comput.* **324**, 202 (2018)
[47] B Kang, Y Deng, K Hewage and R Sadiq, *Int. J. Intell. Syst.* **33(8)**, 1745 (2018)
[48] M Kaiser and C C Hilgetag, *Phys. Rev. E* **69(2)**, 036103 (2004)
[49] M E J Newman, *Proc. Natl. Acad. Sci.* **98(2)**, 404 (2001)
[50] D J Watts and S H Strogatz, *Nature* **393(6684)**, 440 (1998)
[51] M E Newman, *Phys. Rev. Lett.* **89(20)**, 208701 (2002)
[52] L J Allen, *Math. Biosci.* **124(1)**, 83 (1994)
[53] Z Yang and T Zhou, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **85(2)**, 056106 (2012)
[54] W Yan, T Zhou, J Wang, Z Fu and B Wang, *Chin. Phys. Lett.* **22(2)**, 510 (2004)
[55] Z Wang, M A Andrews, Z-X Wu, L Wang and C T Bauch, *Phys. Life Rev.* **15**, 1 (2015)
[56] Z Wang, C T Bauch, S Bhattacharyya, A d'Onofrio, P Manfredi, M Perc, N Perra, M Salathé and D Zhao, *Phys. Rep.* **664**, 1 (2016)
[57] L Lü, D Chen, X-L Ren, Q-M Zhang, Y-C Zhang and T Zhou, *Phys. Rep.* **650**, 1 (2016)
[58] Z Wang, Y Moreno, S Boccaletti and M Perc, *Chaos Solitons Fractals* **103**, 177 (2017)
[59] F Rakowski, M Gruziel, Ł Bieniasz-Krzywiec and J P Radomski, *Physica A* **389(16)**, 3149 (2010)
[60] R W Floyd, *Commun. ACM* **5(6)**, 345 (1962)