International Journal of Innovative
Computing, Information and Control
Volume 12, Number 2, April 2016

ICIC International ©2016 ISSN 1349-4198
pp. 581–590

IDENTIFY INFLUENTIAL SPREADERS IN COMPLEX NETWORKS BASED ON POTENTIAL EDGE WEIGHTS

Jiadong Ren[1,2], Chunyan Wang[1,2,*], Qingshan Liu[1,2], Gang Wang[1,2]
and Jun Dong[1,2]

[1]College of Information Science and Engineering
Yanshan University
[2]The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province
No. 438, Hebei Ave., Qinhuangdao 066004, P. R. China
{ jdren; dongjun }@ysu.edu.cn; *Corresponding author: yan7616494321@sina.com
{ qsliuysu; daaiwujiang6 }@163.com

Received August 2015; revised December 2015

Abstract. *It is a key issue to identify influential spreaders in understanding dynamics of information diffusion in complex networks. In existing methods, the edges are treated equally, but each edge has underlying importance and it may be different. In this paper, a novel method called evidential k-shell centrality based on potential edge weight is proposed to identify influential spreaders. First of all, we propose an edge weighting method based on Jaccard similarity for constructing the weighted networks. Secondly, the value of modified evidential centrality is calculated by considering real degree distribution. Thirdly, combining modified evidential centrality and the layer of nodes located in networks, a new method is proposed to identify influential spreaders. Then, in order to evaluate the performance of the proposed method, we adopt the susceptible-infected (SI) model to simulate the epidemic spreading process by using the spreading rate and the number of infected nodes in real complex networks. Experiment results verify that our method is effective for detecting the node influence.*

Keywords: Complex network, Influential spreaders, Potential edge weight, Dempster-Shafer theory of evidence, k-shell decomposition

1. **Introduction.** In complex network studies, the spreading processes of epidemic and information have gained increasing attention in the recent years. It is extremely significant for developing efficient methods to either hinder the disease spreading or accelerate the information dissemination. Hence, how to identify influential nodes is a crucial issue in complex networks.

It is well-known that degree centrality, betweenness centrality and closeness centrality [1] are three fundamental centrality measures to identify influential nodes. So far, a variety of centrality measures are proposed to identify influential nodes in complex networks [2,3]. Kitsak et al. [4] proposed k-shell (or k-core) decomposition to identify influential nodes, and they found that the influential nodes are those located in the core of the network. However, the k-shell decomposition assigns many nodes in the same k-shell. Then, Zeng and Zhang [5] proposed an improved measure, named the mixed degree decomposition (MDD) method, where both the residual degree and the exhausted degree are considered. In the latest two years, in [6], a measure named coreness centrality was proposed to quantify the spreading capability of a node. The method is based on the idea that a powerful spreader has more connections to nodes that reside in the core of the network. After that, Gao et al. [7] proposed a local structure centrality (LSC) measure which considers both the number and the topological connections of the neighbors of a node.

For nodes with the same number of neighbors, the one with denser connected neighbors is supposed to be more influential since denser connected neighbors get more chance to influence each other.

Dempster-Shafer evidence theory (D-S evidence theory for short) was first proposed by Dempster [8], and then formed by the further expansion of Shafer [9]. D-S evidence theory has the ability to combine a pair of evidence or belief functions to obtain a new evidence or belief function. Based on the D-S theory, evidential centrality (EVC) [10] measure is raised as a tradeoff between degree and strength of each nodes to derive node importance in weighted network. In [11], evidential semi-local centrality (ESC) measure not only modifies the evidential centrality by considering real degree distribution of network, but also extends semi-local centrality in weighted network. The values of centrality measure for each node are obtained by both these centrality measures, respectively. Then, we can rank the nodes and achieve the order according to these values. It can be seen that the higher the value is, the more influential the node is.

In the majority of networks, the edges are treated equally, but each edge may have potential and different significance in network structure [12,13]. It can be seen that the edges' potential importance should be considered. Thus, when we design the centrality measures to identify influential spreaders for the unweighted networks, it is crucial for taking the edges' potential importance into account. However, we find that evidential centrality only captures the characteristics in the aspect of degree, strength, rather than location of the network. If a highly connected node exists at the periphery of a network, it will have a small effect in the spreading process. Conversely, a less connected decision-maker placed in the core of the network will have more influence on other individuals through a network. It seems that the more influential spreaders tend to be those located within the core of a network. Thus, in order to rank nodes effectively, it is better to design the ranking algorithms which consider the location of node in a network. Here, the k-shell decomposition is used to identify the location of a node in the network. Inspired by those, in this paper, a new centrality measure is proposed to identify influential spreaders based on combining modified evidential centrality with taking degree distribution into account and the layer of nodes located in networks. The value of the new centrality measure for each node is ranked in descending order. The higher the value of centrality measure is, the more influential the node is. Then we adopt the susceptible-infected (SI) model to validate the spreading influence of the nodes ranked by different centrality.

The main contributions of this paper can be summarized as below.

- An edge weighting method is proposed according to Jaccard similarity.
- A new centrality measure called evidential k-shell centrality based on potential edge weight (PEW-EKSC) is proposed to identify influential nodes based on combining modified evidential centrality with taking degree distribution into account and the layer of nodes located in networks.
- The new method is raised as a tradeoff between degree and strength of each node, and the layer of nodes located in network is seen as a factor to identify influential spreaders as well.

The rest of parts are organized as follows. In Section 2, we define the potential edge weight and introduce evidence theory. In Section 3, a new method for identifying influential spreaders is proposed. In Section 4, we present data sets and apply the SI models to evaluate the performance of the proposed method. Finally, some conclusions are summarized in the last section.

2. Preliminaries.

2.1. Definition.

Consider an unweighted and undirected network $G = (V, E)$ with $|V| = N$ nodes and $|E| = M$ edges. $e_{uv} \in E$ represents the connection between node u and node v.

The k-shell decomposition method partitions a network into sub-structures that are directly linked to centrality [14]. This method which can identify the location of a node in a network, is an efficient measure at capturing the spreading ability of a node in a network. The algorithm can be implemented in the following way. First of all, we remove all nodes with degree one, and keep on removing the existing nodes until all nodes' degrees are larger than one. The removed nodes belong to the $k - s = 1$ shell. Next, we repeat the pruning process in the same manner for other nodes until all nodes of the network are removed and assigned to the different k-shell index.

In unweighted networks, the edges are treated equally. Generally speaking, the edges are different if the degree of overlap among the neighbors of the nodes connected by an edge is different. Thus, potential edge weights can be defined in the following way.

Definition 2.1. *(Potential Edge Weight) The potential weight of edge e_{uv}, denoted by ω_{uv}, is defined as*

$$\omega_{uv} = \frac{|N\mathrm{g}_u \bigcap N\mathrm{g}_v|}{|N\mathrm{g}_u \bigcup N\mathrm{g}_v|} \tag{1}$$

where $N\mathrm{g}_u$ is a set which consists of neighbor nodes of node u and itself.

2.2. Dempster-Shafer rule of combination.

In Dempster-Shafer evidence theory, problem domain $\Theta = \{a_1, a_2, \cdots, a_n\}$ is a nonempty set which consists of a finite number of mutually exclusive and exhaustive hypotheses, called the frame of discernment.

Suppose Θ is the frame of discernment, a mass function is mapping $m : 2^\Theta \to [0, 1]$, (2^Θ is the power set of the Θ), which is also called Basic Probability Assignment (BPA), satisfying

$$m(\Phi) = 0 \text{ and } \sum_{A \subset 2^\Theta} m(A) = 1 \tag{2}$$

where Φ is the empty set and A is any element of 2^Θ, and mass $m(A)$ represents how strongly the evidence supports A.

Assuming that masses m_1 and m_2 are both basic probability assignments of Θ, orthogonal sum $m(A)$ is calculated from the two sets of masses m_1 and m_2 in Dempster's rule of combination.

$$m(A) = \frac{1}{1 - K} \sum_{B \bigcap C = A} m_1(B) m_2(C) \tag{3}$$

with

$$K = \sum_{B \bigcap C = \Phi} m_1(B) m_2(C) \tag{4}$$

where A, B and C are elements of 2^Θ.

3. Influential Spreaders Identification by Evidential *k*-Shell Centrality Based on Potential Edge Weight.

In this paper, we mainly consider unweighted networks. In unweighted networks, although the edges are on an equal footing, the potential importance of each edge is different. So we present an edge weighting method according to Jaccard similarity. Then we observe that evidential centrality simply weighs between the degree and strength of each node, rather than the layer of nodes located in network. Thus, based on combining modified evidential centrality with taking degree distribution into account and the layer of nodes located in networks, the influence of the node is identified by a

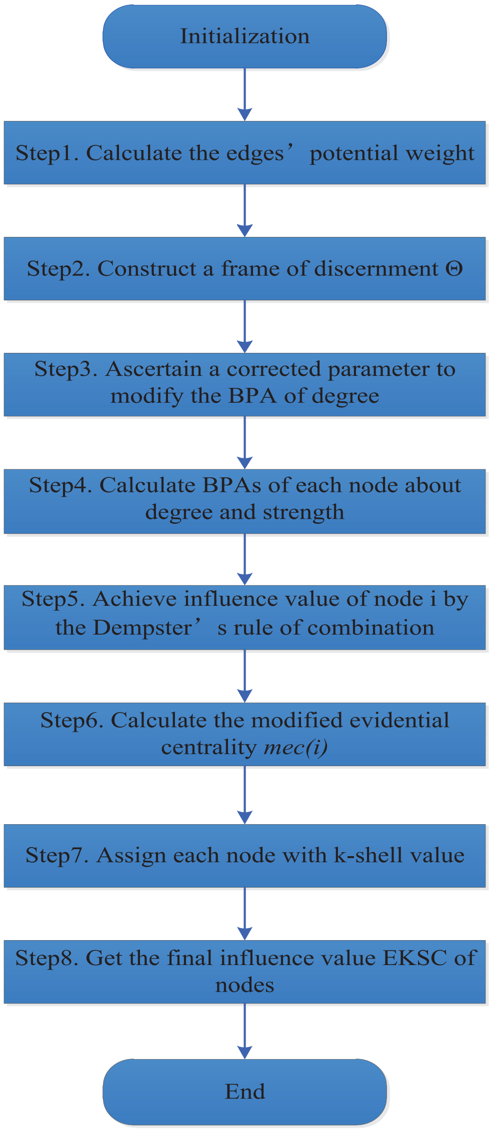


FIGURE 1. Example of figure

new centrality measure, called evidential k-shell centrality based on potential edge weight (PEW-EKSC). The flow chart of the proposed method is shown in Figure 1.

In the following, we analyze the algorithm in detail.

Step 1. Since the potential importance of the edges plays a significant role on identifying influential nodes, the edges' potential weights in unweighted networks are calculated based on the definition of the potential edge weight.

Step 2. In the networks with potential edge weights, to evaluate the influence of degree and strength of node, here, let *high* or *low* be evaluation indices. Hence, a frame of discernment θ is denoted as $\theta = (high, low)$.

Step 3. The real degree distribution of a complex network should be taken into consideration. Thus, suppose node i with degree k_i follows a degree distribution $P(k_i)$. To modify the BPA of degree, the parameter is ascertained and defined as $\lambda_i = \sum_{j \le k_i} P(j)$, where j is a set of degree of nodes which is lower than k_i.

Step 4. The BPAs of *high* or *low* influence for the degree of the ith node are represented for $m_{id}(h)$ or $m_{id}(l)$ $(i = 1, 2, \ldots, N)$, separately; Likewise, BPAs of *high* or *low* influence for the strength of the ith node are represented for $m_{i\omega}(h)$ or $m_{i\omega}(l)$ $(i = 1, 2, \ldots, N)$, respectively. They are expressed as follows.

$$m_{id}(h) = \lambda_i \frac{|k_i - k_m|}{\sigma} \tag{5}$$

$$m_{id}(l) = (1 - \lambda_i)\frac{|k_i - k_M|}{\sigma} \tag{6}$$

$$m_{i\omega}(h) = \frac{|\omega_i - \omega_m|}{\delta} \tag{7}$$

$$m_{i\omega}(l) = \frac{|\omega_i - \omega_M|}{\delta} \tag{8}$$

where σ and δ are given as

$$\sigma = k_M + \mu - (k_m - \mu) = k_M - k_m + 2\mu \tag{9}$$

$$\delta = \omega_M + \varepsilon - (\omega_m - \varepsilon) = \omega_M - \omega_m + 2\varepsilon \tag{10}$$

where $0 < \mu < 1$, $0 < \varepsilon < 1$, [10] demonstrated that the values of μ and ε have no impact on the ranking orders of nodes in weighted network. k_M and k_m are the maximum and minimum values of degree, and ω_M and ω_m correspond to the maximum and minimum values of weight, respectively. According to the above statement, the BPAs of degree and strength of ith node are obtained, respectively, as

$$M_d(i) = (m_{id}(h), m_{id}(l), m_{id}(\theta)) \tag{11}$$

$$M_\omega(i) = (m_{i\omega}(h), m_{i\omega}(l), m_{i\omega}(\theta)) \tag{12}$$

where

$$m_{id}(\theta) = 1 - (m_{id}(h) + m_{id}(l)) \tag{13}$$

$$m_{i\omega}(\theta) = 1 - (m_{i\omega}(h) + m_{i\omega}(l)) \tag{14}$$

where $m_{id}(\theta)$ represents the BPAs of high or low influence for the degree of the ith node and $m_{i\omega}(\theta)$ represents the BPAs of high or low influence for the strength of the ith node.

Step 5. For the above BPAs of the ith node with respect to degree and strength, the BPA of influence value of the ith node is achieved by the Dempster's rule of combination, and is given by

$$M(i) = (m_i(h), m_i(l), m_i(\theta)) \tag{15}$$

where $\theta = (high, low)$, and $m_i(h)$ and $m_i(l)$ represent the BPAs of high or low influence of the ith node after using the Dempster's rule of combination, respectively. In Equation (15), $m_i(\theta)$ means the probability of *high* or *low* of the ith node. Thus, let $m_i(\theta)$ allocate to $m_i(h)$ and $m_i(l)$ normally, then the probabilities of *high* or *low* influence of the ith node are given by

$$M_i(h) = m_i(h) + \frac{1}{2m_i(\theta)} \tag{16}$$

$$M_i(l) = m_i(l) + \frac{1}{2m_i(\theta)} \tag{17}$$

Step 6. Apparently, the higher the value of $M_i(h)$ is, the more important the node i is. In contrast, the lower the value of $M_i(l)$ is, the stronger the spreading influence of node i is. So the modified evidential centrality $mec(i)$ of the ith node is defined as

$$mec(i) = M_i(h) - M_i(l) = m_i(h) - m_i(l) \tag{18}$$

In Equation (18), the value of $mec(i)$ is a positive or negative number. Thus, to ensure $mec(i)$ to be a positive number, the numerical treatment and normalization are denoted as below.

$$mec(i) = \frac{|\min(mec)| + mec(i)}{\sum_{i=1}^{N}\{|\min(mec)| + mec(i)\}} \tag{19}$$

where $|\min(mec)|$ is the absolute minimum value of mec.

Step 7. Assign each node with k-shell value by employing the k-shell decomposition analysis. Lastly, normalization is necessary.

$$K_s(i) = \frac{K_s(i)}{\sum_i K_s(i)} \tag{20}$$

Step 8. The final influence value EKSC for each node can be achieved according to the expression as below.

$$EKSC(i) = mec(i) \times K_s(i) \tag{21}$$

Finally, the value of EKSC for each node is ranked in descending order. Thus, nodes located in the core of networks have larger EKSC(i) than ones in the periphery of a network. To sum up, the higher the value of EKSC(i) is, the more influential the node is.

4. **Experimental Analysis.** In this section, to validate the performance of the proposed method, we employ the susceptible-infected (SI) model [14] to examine the spreading influence of top-ranked nodes by different methods. In the SI model, there are two compartments: (i) *Susceptible* $S(t)$ represents the number of individuals susceptible to (not yet infected) the disease; (ii) *Infected* $I(t)$ denotes the number of individuals that have been infected and are able to spread the disease to susceptible individuals. At the initial time, we set only a node to be infected to investigate the influence of the node. At each step, each node in the infected state randomly selects their susceptible neighbors with probability β and remains infected. The spreading process stops when there is no infected node. The number of infected nodes at time t, denoted by $F(t)$, can be treated as an indicator to estimate the influence of initially infected node. The larger $F(t)$ value of a node is, the stronger the spreading influence of the node is. The results are obtained by averaging over 500 independent realizations.

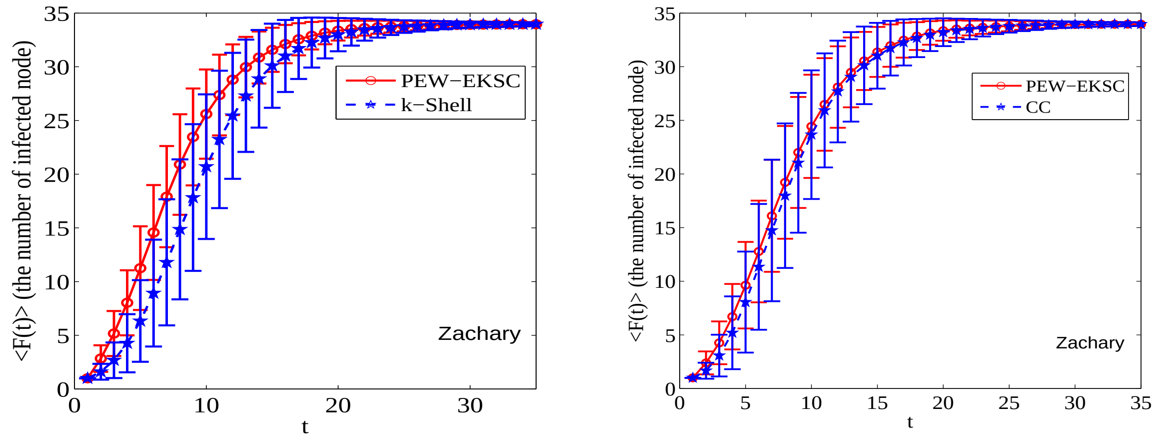
4.1. **Experimental data.** We evaluate the performance of the proposed method in four real networks. (i) Zachary's Karate Club Network [15], the undirected and unweighted network consists of 34 nodes. The data is collected from the members of a university karate club by Wayne Zachary. (ii) Jazz Musicians Network [16], this is the collaboration network between Jazz musicians which has 198 nodes. (iii) Dolphins Network [17], this is an undirected social network of bottlenose dolphins. (iv) Email Network [18]: a network of e-mail interchanges between members of the University Rovira i Virgili (Tarragona). The basic topological properties of these four networks including the number of nodes n

Table 1. The basic topological features of the four real networks

Network	n	m	$< k >$	$< k^2 >$	β^c_{rand}
Zachary	34	78	4.5882	35.6471	0.1287
Jazz	198	2742	37.6970	1070.2424	0.0259
Dolphin	62	159	5.1290	34.9032	0.1470
Email	1133	5451	9.6222	179.1640	0.0535

and links m, the average degree $< k >$, the second-order average degree $< k^2 >$, and the spreading threshold ($\beta^c_{rand} = \frac{<k>}{<k^2>}$) [19] are shown in Table 1.

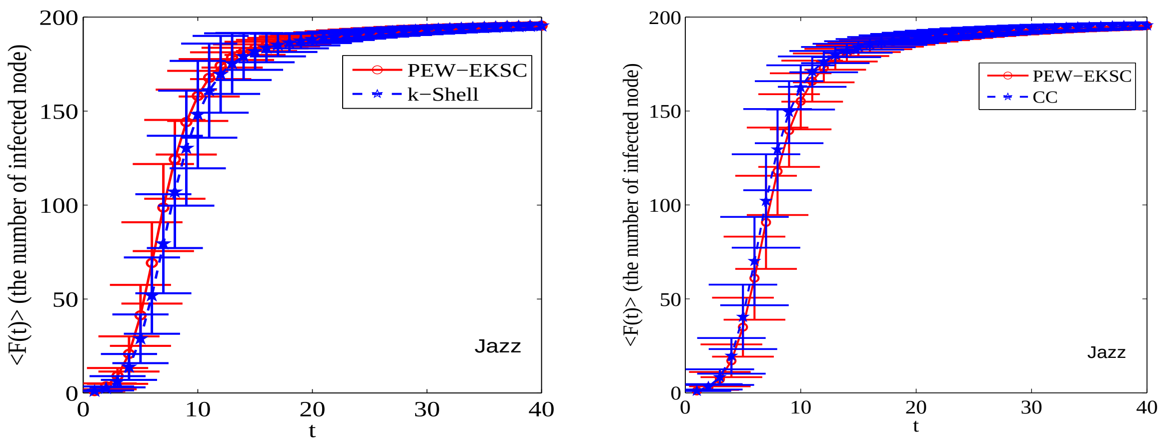
4.2. **Experimental results.** To verify the effectiveness of the proposed method, we compare the spreading ability of the nodes that either appear in the top-L list by PEW-EKSC or other two centrality measures including k-shell decomposition and closeness centrality (not appearing in both lists). Note that without considering the effects of common nodes in both ranking lists, the differences of these methods can be well distinguished. In this paper, we are more interested in the area when β is around the epidemic threshold β^c_{rand} since in this area the initially infected nodes are perturbations that can trigger spreading at all scales. Thus, we use SI model to compare the spreading ability of the nodes on four real networks. The simulations on the cumulative number of infected nodes, namely $F(t)$,



(a) Comparison by PEW-EKSC or k-Shell decomposition

(b) Comparison by PEW-EKSC or CC

FIGURE 2. The number of infected nodes by initially infected nodes in the top-5 list



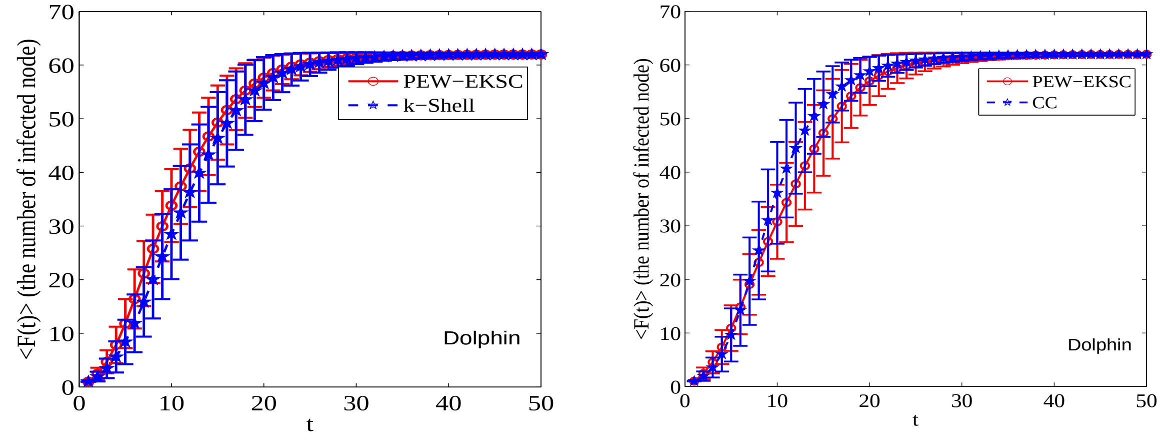
(a) Comparison by PEW-EKSC or k-Shell decomposition

(b) Comparison by PEW-EKSC or CC

FIGURE 3. The number of infected nodes by initially infected nodes in the top-20 list

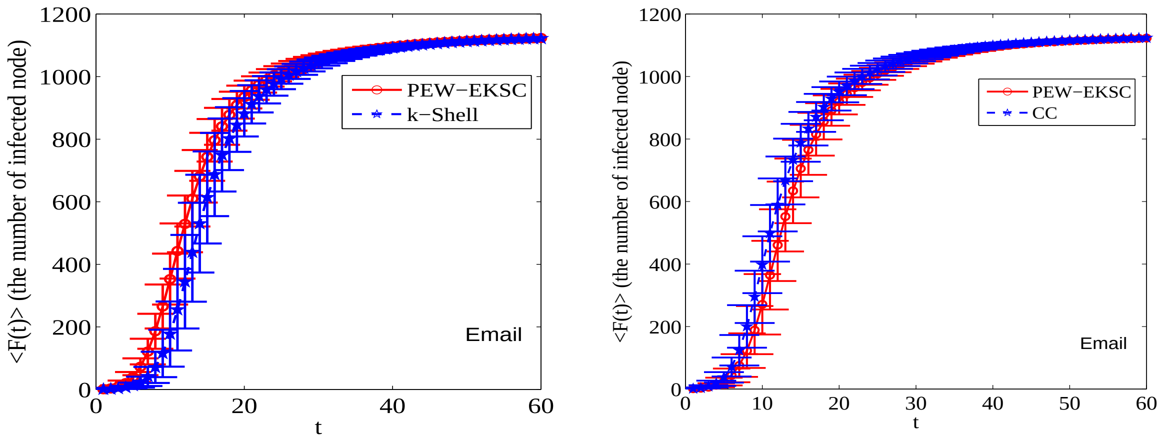
as a function of time for Zachary and Jazz networks are shown in Figure 2 and Figure 3, respectively, and those for Dolphin and Email network are shown in Figure 4 and Figure 5, separately. The number of cumulative infected nodes increases with time and finally reaches the steady value.

As shown in Figure 2(a) and Figure 2(b), we can find that the proposed method is better than CC for the average number of infected nodes at each step during the whole infecting process. Obviously, the proposed method outperforms k-shell decomposition method in Zachary Club network. In Jazz network, Figure 3(a) presents the cumulative number of infected nodes by the proposed method is larger than that by k-shell decomposition method in each step, that is the proposed method performs a quicker spreading than k-shell decomposition method, but the average number of infected nodes by the proposed



(a) Comparison by PEW-EKSC or k-Shell decomposition (b) Comparison by PEW-EKSC or CC

FIGURE 4. The number of infected nodes by initially infected nodes in the top-10 list



(a) Comparison by PEW-EKSC or k-Shell decomposition (b) Comparison by PEW-EKSC or CC

FIGURE 5. The number of infected nodes by initially infected nodes in the top-10 list

method and CC is almost the same as shown in Figure 3(b), it means that the proposed method and CC nearly have the same performance on Jazz. In Dolphin network, it is obviously that the result for the proposed method is a bit better than the result for k-shell decomposition method in Figure 4(a). However, comparing the proposed method with CC, as shown in Figure 4(b), in early stage, the spreading speed of the top-L nodes ranked by proposed method is faster than that by the CC, but the CC needs less step to reach the steady state than the proposed method in the later spreading process. Furthermore, when considering the Email network, the results for the comparison of the spreading ability as to proposed method, k-shell decomposition method and CC are shown in Figure 5(a) and Figure 5(b). The proposed method performs completely better than k-shell decomposition method, but CC outperforms the proposed method to some extent.

Totally speaking, from the comparison with the proposed method and other centrality measures, our method is effective for identifying the influential spreaders in real networks.

5. **Conclusions.** In this paper, we propose a new approach to identify influential spreaders in complex network called evidential k-shell centrality based on potential edge weight. Firstly, an edge weighting method is proposed according to Jaccard similarity for constructing weighted networks. Secondly, the value of modified evidential centrality is calculated by considering real degree distribution. Thirdly, combining modified evidential centrality, the layer of nodes located in network is seen as a factor to identify influential spreader as well. To evaluate the performance of our method, we use the susceptible-infected (SI) model to estimate spreading influence of the top-L nodes in four real networks. By comparing the number of cumulative infected nodes of the top-L spreaders by different centrality measures, it shows that our method performs a quicker spreading than k-shell decomposition method and CC in four networks. The experiment results on four real networks show that our method is capable of identifying influential spreaders well. Further investigation could be done on identifying influential spreaders in directed networks. Also the potential edge weight might be a useful factor in estimating the spreading ability of a node in unweighted networks.

Acknowledgment. This work is supported by the National Natural Science Foundation of China under Grant No. 61572420, No. 61472341 and the Natural Science Foundation of Hebei Province China under Grant No. F2013203324, No. F2014203152 and No. F2015203326.

REFERENCES

[1] L. Freeman, Centrality in social networks conceptual clarification, *Social Networks*, pp.215-239, 1979.
[2] J. Borge-Holthoefer and Y. Moreno, Absence of influential spreaders in rumor dynamics, *Physical Review E*, vol.85, no.2, 2012.
[3] J. Borge-Holthoefer, A. Rivero and Y. Moreno, Locating privileged spreaders on an online social network, *Physical Review E*, vol.85, no.6, 2012.
[4] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley and H. A. Makse, Identification of influential spreaders in complex networks, *Nature Physics*, vol.6, pp.888-893, 2010.
[5] A. Zeng and C. J. Zhang, Ranking spreaders by decomposing complex networks, *Physics Letters A*, vol.377, pp.1031-1035, 2013.
[6] J. Bae and S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, *Physica A*, vol.395, pp.549-559, 2014.
[7] S. Gao, J. Ma, Z. M. Chen, G. H. Wang and C. M. Xing, Ranking the spreading ability of nodes in complex networks based on local structure, *Physica A*, vol.403, pp.130-147, 2014.
[8] A. P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, *Annals of Mathematics and Statistics*, vol.38, no.2, pp.325-339, 1967.
[9] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.

[10] D. Wei, X. Deng, X. Zhang, Y. Deng and S. Mahadevan, Identifying influential nodes in weighted networks based on evidence theory, *Physica A*, vol.392, pp.2564-2575, 2013.
[11] C. Gao, D. Wei, Y. Hu, S. Mahadevan and Y. Deng, A modified evidential methodology of identifying influential nodes in weighted networks, *Physica A*, vol.392, pp.5490-5500, 2013.
[12] Z. He, S. Liu and M. Zhan, Dynamical robustness analysis of weighted complex networks, *Physica A*, vol.392, pp.4181-4191, 2013.
[13] C. J. Zhang and A. Zeng, Network skeleton for synchronization: Identifying redundant connections, *Physica A*, vol.402, pp.180-185, 2014.
[14] T. Zhou, J. G. Liu, W. J. Bai, G. Chen and B. H. Wang, Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity, *Physical Review E*, vol.74, 2006.
[15] W. Zachary, An information flow model for conflict and fission in small groups, *Journal of Anthropological Research*, vol.33, 1977.
[16] P. Gleiser and L. Danon, Community structure in Jazz, *Adv. Complex Syst.*, vol.6, 2003.
[17] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten and S. M. Dawson, The bottlenose Dolphin community of doubtful sound features a large proportion of long-lasting associations, *Behav. Ecol. Sociobiol.*, vol.54, pp.396-405, 2003.
[18] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt and A. Arenas, Self-similar community structure in a network of human interactions, *Physical Review E*, vol.68, 2003.
[19] C. Castellano and R. Pastor-Satorras, Thresholds for epidemic spreading in networks, *Phys. Rev. Lett.*, vol.105, 2010.