Identifying Vital Nodes in Social Networks Using an Evidential Methodology Combining with High-Order Analysis

Meng Zhang, Guanghui Yan(✉), Yishu Wang, and Ye Lv

School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
736336430@qq.com

Abstract. Identifying vital nodes is a basic problem in social network research. The existing theoretical framework mainly focuses on the lower-order structure of node-based and edge-based relations and often ignores important factors such as interactivity and transitivity between multiple nodes. To identify the vital nodes more accurately, a high-order structure, named as the motif, is introduced in this paper as the basic unit to evaluate the similarity among the node in the complex network. It proposes a notion of high-order degree of nodes in complex network and fused the effect of the high-order structure and the lower-order structure of nodes, using evidence theory to determine the vital nodes more efficiently and accurately. The algorithm was evaluated from the function of network structure. And the SIR model was adopted to examine the spreading influence of the nodes ranked. The results of experiments in different datasets demonstrate that the algorithm designed can identify vital nodes in the social network accurately.

Keywords: Vital nodes · High-order network · Evidence theory · SIR

1 Overview

In pace with the rapid development of information technology, the forms of communication and interaction have diversified. The resulting massive data can not only help us better understand the relationship between people, but also show the mode of information transmission between people [1–3]. Identifying vital nodes in the network helps us to guide the information dissemination better.

Centrality is a method which can measure the significance or importance of actors in social networks. Considerable centrality measures have been carried out previously for ranking the nodes based on network topology such as Degree Centrality (DC) [4], Closeness Centrality (CC) [5] and Betweenness Centrality

Supported by the Natural Science Foundation of China (No. 61662066, 61163010).

J. Zeng et al. (Eds.): ICPCSEE 2020, CCIS 1257, pp. 101–117, 2020.
https://doi.org/10.1007/978-981-15-7981-3_8

(BC) [6,7]. Although the DC is intuitive and simple, we just take into consideration the degree of nodes and ignores the global structure of the network. Whether a node is important affects the importance of its neighbor nodes, which in the social network its neighbor nodes would follow the behavior of the former. The other two metrics based on the global structure of a network can better characterize the importance of nodes, but exhibit some serious drawbacks because of the computational complexity in large-scale networks [8,9]. Considering the operating efficiency and experimental results of the algorithm, Chen et al. [8] proposed an effective Semi-local Centrality. These above methods take nodes and edges as research objects. Despite the success of these methods in identifying vital nodes, they ignore the possible relationship between nodes, which may lead us to deviate from the overall cognition of the network. So, an important issue is that how to describe the interaction, transitivity and other factors between nodes.

A number of researches have shown that social networks contain abundant subgraph structures, which are characterized by transitivity, interaction and so on [10,11]. Usually, we describe this subgraph structure as network motif or graphlet [11–13]. Compared with the method of researching from edges and points, the network structure with small subgraph structure as the research unit is called high-order network structure. In broader network analysis, high-order structure is often described through the idea of a network motif. Since the concept of motif was put forward in 2002 by Moli R et al., most of the research had centered on how to count the number of motif efficiently in the network [10,14,15]. Until 2016, Benson et al. proved that motif can be used for graph clustering and community discovery, and proposed a series of theoretical basis. The research of high-order network has become one of the important means of current research [10,15].

The importance of nodes in the network is a vague and relative concept. As it happens, the Dempster–Shafer evidence theory is a complete theory for dealing with uncertainty information. It was first proposed by Dempster and then perfected by Shafer. Compared with traditional probability theory, D-S evidence theory can not only express random uncertainty, but also express incomplete information and subjective uncertainty information [8,16,17]. D-S evidence theory also provides a powerful Dempster combination rule for information fusion, which can achieve the fusion of evidence without prior information, and can effectively reduce the uncertainty of the system. On this basis, Wei et al. put forward an algorithm to rank nodes in weighted networks [9].

In this paper, based on the high-order network analysis and Dempster-Shafer evidence theory, we designed a high-order evidence semi-local centrality to identify vital users accurately in the social network. First, we designed a concept based on high-order structure. And then the low-order information and high-order information of nodes are regarded as two basic probability assignments (BPAs). On the one hand, we verified the rationality of the proposed method through the function of network structure. On the other hand, we adopted the Susceptible-Infected-Recover (SIR) model to examine the spreading influence of the top nodes by different centrality measures. The experiments on real social networks are applied to show the accuracy of the proposed method.

2 Related Work

High-order thinking has been shown to be useful in many applications such as social networks, biology, neuroscience, and so on [15]. Network motifs are the basic building blocks of networks and are also one of the important expressions of high-order network structure [10,11].

2.1 High-Order Network Structure

Combined with the basic theory of sociology, this paper takes the 3-order motif as the basic. Figure 1 shows the 13 connection modes of the 3-order motif.

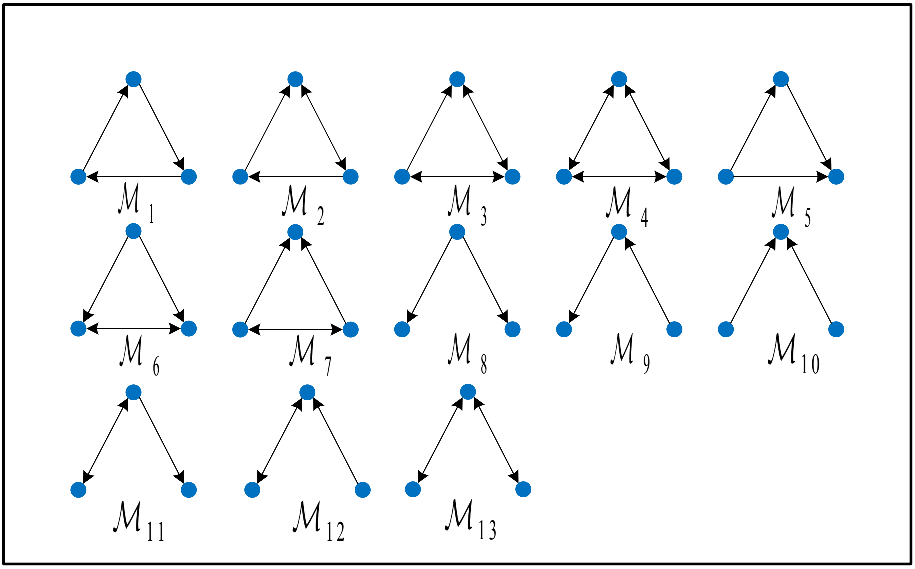


Fig. 1. All 13 connection modes of the 3-order motif $\mathcal{M}$ [15]

A motif $\mathcal{M}$ is usually defined as a tuple $(B, \mathcal{A})$ on k nodes, where B is a $k \times k$ binary matrix and $\mathcal{A} \subset \{1, 2, \cdots, k\}$ is a set of anchor nodes [10].

$$\mathcal{M}\left(B, \mathcal{A}\right) = \left\{\left(set\left(\mathbf{v}\right), set\left(\mathcal{X}_{\mathcal{A}}\left(\mathbf{v}\right)\right)\right) | \mathbf{v} \in V^k, v_1, \cdots, v_k, distinct, \mathbf{A}_{\mathbf{v}} = B\right\} \tag{1}$$

Given a motif $\mathcal{M}$, we define the motif-based adjacency matrix by $\mathbf{A}_{\mathcal{M}} = \{a_{ij}\}_{N \times N}$. In this paper, we use Benson et al. [11] improved algorithm to get the motif-based adjacency matrix. So, the algorithm is as follows [11,18]:

Algorithm 1 Algorithm for the motif-based adjacency matrix $\mathbf{A}_{\mathcal{M}}$.

Input: Directed network $G = (V, E)$ and selected motif $\mathcal{M}$
Output: $\mathbf{A}_{\mathcal{M}}$;

1: pro–processing: If $\mathcal{M}$ is $\mathcal{M}_4$, ignore all undirectional edges in G. If $\mathcal{M}$ is $\mathcal{M}_1$ or $\mathcal{M}_5$, ignore all bidirectional edges in G.
2: Obtain the undirected graph G_1 by getting rid of the direction of all edges in G.
3: d_u is the degree of node v_i in G_1. Sort the nodes in G_1 by ascending degree.
4: For every edge undirected edge u,v in G_1, if $d_u < d_v$, add directed edge (u,v) to G_2; otherwise, add directed edge (v,u) to G_2.
5: For every node in u in G_2 and every pair of directed edges (u,v) and (u,w), check to see if edge (v,w) or (w,v) is in G_2. If so, check whether these three nodes form motif $\mathcal{M}$ in G. If they do, increment the weights of edges $(\mathbf{A}_{\mathcal{M}})_{uv}$, $(\mathbf{A}_{\mathcal{M}})_{uw}$, and $(\mathbf{A}_{\mathcal{M}})_{uv}$ by 1.
return $\mathbf{A}_{\mathcal{M}}$ as the motif-based adjacency matrix;

Let $G = (V, E, \mathbf{A}_{\mathcal{M}})$ be a directed and unweighted graph, where $|V| = \{v_i | i = 1, 2, 3, \ldots, n\}$ is the node set, and $|E| = \{e_{ij} | i, j = 1, 2, \ldots, n\}$ is the arc set, where e_{ij} is a directed edge from v_i to v_j. $\mathbf{A}_{\mathcal{M}}$ is the motif-based adjacency matrix.

2.2 Centrality Measures

Roughly speaking, there are two kinds of method about identifying vital nodes which are based on the number of neighbor and based on path in network. The former is characterized by the degree of nodes in the network as a measure of importance. This method is relatively intuitive and has good performance, such as DC. The latter measures the importance of nodes by controlling the information flow in the network, such as BC and CC. This kind of method is relatively complex and not suitable for large-scale networks [20].

The semi-local centrality is a node importance ranking method based on the number of neighbor in the networks. The method not only considered the neighbors of the nodes but also the neighbors and next neighbors of the neighbors. In other words, this method has low time complexity and is suitable for large-scale networks. Semi-local centrality of node v_i is defined as [9]:

$$SLC(i) = \sum_{j \in \Gamma(i)} \sum_{k \in \Gamma(j)} N^w(k) \tag{2}$$

where $N^w(k)$ is the number of the nearest and the nextest neighbors of node v_k, and $\Gamma(i)$ is the set of the nearest neighbors of node v_i.

2.3 Dempster-Shafer Theory of Evidence

The essence of D-S evidence theory is a generalization of probability theory. The basic event space in probability theory is extended into the power set space of the basic event, and the basic probability assignment function is established on it [13,20].

Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$ be a finite complete set of N elements which is mutually exclusive. The frame of discernment is the set Θ. The power set of Θ is denoted as 2^Θ which is composed of 2^N elements.

$$2^\Theta = \{\varnothing, \theta_1, \theta_2, \ldots \theta_N, \theta_1 \cup \theta_2, \ldots, \theta_1 \cup \theta_2 \cup \theta_3, \ldots, \Theta\} \tag{3}$$

For a frame of discernment Θ, a basic probability assignment function is a mapping $m : 2^\Theta \rightarrow [0,\ 1]$, satisfying two conditions as follows:

$$m\left(\varnothing\right) = 0 \tag{4}$$

and

$$\sum_{A \subseteq \Theta} m\left(A\right) = 1 \tag{5}$$

where m is called the basic probability assignment (BPA), and $m\left(A\right)$ represents how strongly the evidence supports A.

In order to combine with information from multiple independent information sources, D-S evidence theory provides Dempster's Rules of Combination to achieve the fusion of multiple evidence. Its essence is the orthogonal sum of evidence.

$$\begin{cases} m\left(\varnothing\right) = 0 \\ m\left(A\right) = \frac{1}{1-k} \sum_{A_i \cap B_i = A} m_1\left(A_i\right) m_2\left(B_i\right) \end{cases} \tag{6}$$

where k is a normalization constant, called the conflict coefficient of BPAs.

$$k = \sum_{A_i \cap B_i = \varnothing} m_1\left(A_i\right) m_2\left(B_i\right) \tag{7}$$

3 High-Order Evidential Semi-local Centrality

In this section, some notations and some knowledge of high-order degree are given as follows. Two BPAs of a node are obtained Eq. 9 based on the high-order degree and degree of the node, respectively. An evaluation method of importance of the node is established by Dempster's rule of combination [9]. The influence of the node is identified by a new centrality measure, called the high-order evidential centrality. Inspired by semi-local centrality [English16], we propose high-order evidential semi centrality since it not only fuse the internal information of the network, but also considers the global structure information.

3.1 High-Order Degree

In the high-order network structure, the elements of motif-based adjacency matrix describe the local connection density of node pair (v_i, v_j). The higher the weight, the more the modal structures with the edge of node pairs are, the worse the anti-attack ability of the node pairs is, the higher the importance of the node pairs is. The higher the value of elements is, the more the number of motifs which contain the edge of node pairs are, that represent the worse the anti-attack ability of the node pair is, the higher the importance of the node pairs is [20].

The high-order degree of node v_i is the number of times node v_i appears in the given motif. The high-order degree of the node v_i denote as H_i. The H_i algorithm is as follows:

Algorithm 2 Algorithm for the H_i.

Input: Undirected network G_1, directed network G_2 and selected motif $\mathcal{M}$
Output: The number of $\mathcal{M}$ and high-order H_i

1: initialize: the H_i of all nodes in G_1.
2: for edge (u, v) node v_w in G_1 do:
3: if node v_w is not node v_v:
4: if edge (v, w) in G_2:
5: $\mathcal{M}$ which consist of node v_u, node v_v and node v_w is isomorphic with subgraph which consist of node v_u, node v_v and node v_w in G_1
6: the value of H_v, H_u and H_w plus one respectively
7: end

3.2 BPAs of Degree and High-Order Degree

A node in the network is either important or not important, so we ascertain a frame of discernment Θ about each node, so a frame of discernment Θ is given as [9]:

$$\Theta = \{h, l\} \tag{8}$$

where h and l represent important and unimportant respectively which are two mutually exclusive elements.

The degree and the high-order degree are two indicators of importance about each node. And then we can obtain these two basic probability assignment functions from different independent sources.

So, two basic probability assignment functions are given as follows:

$$\begin{array}{l} m_{d_i} : m_{d_i}(h), m_{d_i}(l), m_{d_i}(\theta) \\ m_{H_i} : m_{H_i}(h), m_{H_i}(l), m_{H_i}(\theta) \end{array} \tag{9}$$

where $m_{d_i}(\theta)$ and $m_{H_i}(\theta)$ represent the probability whether a node is important or not in the above two indicators. And their value are

$$\begin{array}{l} m_{d_i}(\theta) = 1 - \left(m_{d_i}(h) + m_{d_i}(l)\right) \\ m_{H_i}(\theta) = 1 - \left(m_{H_i}(h) + m_{H_i}(l)\right) \end{array} \tag{10}$$

$$\begin{aligned} m_{d_i}(h) &= \lambda_i \tfrac{|k_i - k_m|}{\sigma} \\ m_{d_i}(l) &= (1-\lambda_i) \tfrac{|k_i - k_M|}{\sigma} \\ m_{H_i}(h) &= \tfrac{|H_i - H_m|}{\delta} \\ m_{H_i}(l) &= \tfrac{|H_i - H_M|}{\delta} \end{aligned} \tag{11}$$

where σ and δ are given as:

$$\begin{aligned} \sigma &= k_M + \mu - (k_m - \mu) = k_M - k_m + 2\mu \\ \delta &= w_M + \epsilon - \left(w_m - \epsilon\right) = w_M - w_m + 2\epsilon \end{aligned} \tag{12}$$

μ and ϵ are given as 0.15. Because the value of μ and ϵ have no effect on the results. The influence value of node v_i is obtained by Dempster–Shafer theory of evidence, and is given by [9]:

$$M(i) = (m_i(h), m_i(l), m_i(\theta)) \tag{13}$$

Normally, let $m_i(\theta)$ assign to $m_i(h)$ and $m_i(l)$ averagely, then

$$\begin{aligned} M_i(h) &= m_i(h) + \tfrac{1}{2m_i(\theta)} \\ M_i(l) &= m_i(l) + \tfrac{1}{2m_i(\theta)} \end{aligned} \tag{14}$$

where $M_i(h)$ and $M_i(l)$ are the probability of importance and unimportance about node v_i, respectively. For node v_i, the higher the value of $M_i(h)$ is, the more important the node is. In other words, the lower the value of $M_i(l)$ is, the less important the node is [7,21].

The high-order evidential centrality $hec(i)$ of node v_i is defined as

$$hec(i) = M_i(h) - M_i(l) = m_i(h) - m_i(l) \tag{15}$$

where $hec(i)$ is a positive or negative number. In order to ensure $hec(i)$ is a positive number. The numerical treatment and normalization are denoted as follows,

$$HEC(i) = \frac{|\min(hec)| + hec(i)}{\sum_{i=1}^{N} \{|\min(hec)| + hec(i)\}} \tag{16}$$

The example is a directed network with 10 nodes, see Fig. 2, and k_i, H_i, $m_i(h)$, $m_i(h)$ and $HEC(i)$ for a single node is listed in Table 1.

3.3 High-Order Evidential Semi-local Centrality

We can calculate the value of HEC about each node though the above measure. Inspired by the semi-local centrality measure, we use HEC instead of degree of each node and then high-order evidential semi-local centrality (HESC) is defined. The HESC algorithm is as follows:

$$Q(j) = \sum_{k \epsilon \Gamma(j)} N^w(k) \tag{17}$$

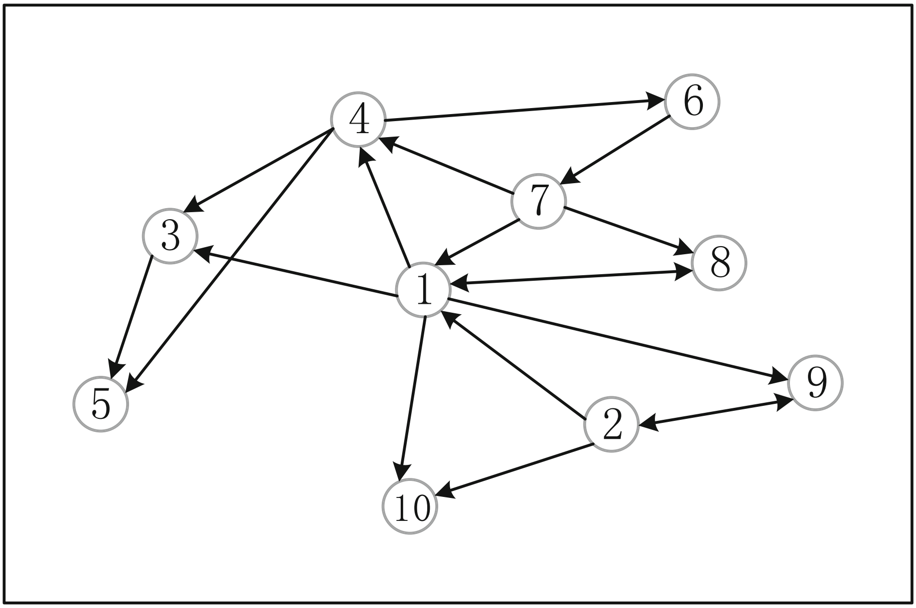


Fig. 2. High-order structures in network motifs $\mathcal{M}_{\triangledown}$

Table 1. An example of a HEC

Nodes	k_i	H_i	$m_i(h)$	$m_i(h)$	$HEC(i)$
1	5	3	0.9177	0	0.3033
2	3	1	0.2918	0.6452	0.0957
3	1	2	0.4817	0.4573	0.1574
4	3	3	0.9007	0.0316	0.2954
5	0	1	0.0964	0.8747	0.0263
6	1	0	0.0074	0.9466	0
7	3	1	0.2918	0.6452	0.0957
8	1	0	0.0074	0.9466	0
9	1	0	0.0074	0.9466	0
10	0	1	0.0964	0.8747	0.0263

$$HESC(i) = \sum_{j \epsilon \Gamma(i)} Q(j) \tag{18}$$

where $N^w(k)$ is the sum of HEC of nearest and next nearest neighbors of node v_k, and $\Gamma(i)$ is the set of the nearest neighbors of node v_i.

4 Example and Experimental Analysis

In this section, we will use the proposed HESC to obtain the ranking of nodes in three different social networks. Meanwhile, comparing with another four traditional centrality measures (DC, CC, BC and EC), we will show the difference between them.

4.1 Datum

We conducted experiments on three social network datasets. Specific Description about the three datasets is shown below:

Advogato. This is the trust network of Advogato. Advogato is an online community platform for developers of free software launched in 1999. Nodes represent Advogato users and a directed edge represent trust relationships called "certification". Advogato have three levels of "certification" corresponding to three edge weights, the weight of master is 1.0, the weight of journeyer is 0.8 and the weight of apprentice is 0.6. An observer without any trust certifications can only trust himself, and therefore the network contains loops [22].

Wiki-Vote. This is a social network that describes voting relationships among Wikipedia users. The network contains all the Wikipedia voting relationships from the inception of Wikipedia to January 2008. In this network, nodes can be regard as Wikipedia users and a directed edge from node v_i to node v_j is that user v_i voted on user v_j [23].

Table 2. Dataset statistics about three social networks

Dataset	Nodes	Edges	The number of $\mathcal{M}_1$ to $\mathcal{M}_7$
Advogato	6.5K	51.1K	18.3K
Wiki-Vote	7.1K	103.7K	608.4K
soc-Epinions1	75.9K	508.8K	1.6M

soc-Epinions1. This data set describes a who-trust-whom online social network of a general consumer review site Epinions.com. Whether to trust each other is decided by the members of the site. Reviews are shown to user based on the Web of Trust and review ratings. Nodes in the network represent consumers, and the directed edges are the trust relationship between consumers [24] (Table 2).

4.2 Relation Between Centrality Measures

In order to more intuitively characterize the relationship between different centrality measures, we compared the average centrality value of each node in the network by averaging over 100 independent runs under different centralities. The relationship between HESC and the other five centrality measures is shown in Fig. 4, Fig. 3 and Fig. 5, respectively. The Y-axis is the value of HESC. And the X-axis is the value of other centrality measures. From the Fig. 4, Fig. 3 and Fig. 5 we can see that the correlation between HESC and DC are the strongest as positively correlated (i.e., Fig. 4(a), Fig. 3(a) and Fig. 5(a)). And then because the HESC value of each node is obtained through the HODC value of each node and the local information of the network is considered at the same time, the correlation between HESC and DC is positive (Table 3).

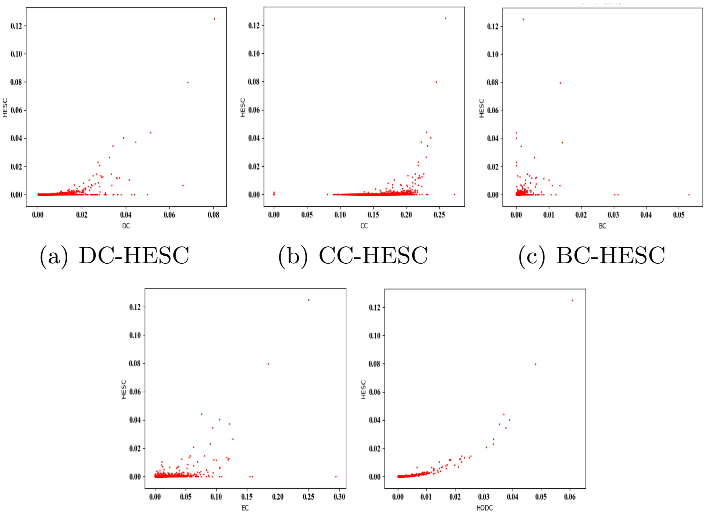
4.3 Experimental Results Analysis

We remove the vital nodes in the network in turn, and compare the relative size of strong connected subgraphs to judge the network invulnerability under static attack. In this experiment, according to the order of above different centrality methods, 10 nodes are removed each time, and then the relative size of strongly connected subgraphs is calculated. Figure 6 shows the change of the relative size of the strong connected subgraphs corresponding to the five methods in different datum when removing the topn nodes from the network under static attack. X-axis represents the number of nodes removed from the network in order, and Y-axis represents the relative size of strongly connected component of the network.

Table 3. The number of various motifs in three datasets

$\mathcal{M}$	Advogato	Wiki-Vote	soc-Epinions1
$\mathcal{M}_1$	63	6795	7656
$\mathcal{M}_2$	2230	17667	84384
$\mathcal{M}_3$	3162	15275	328076
$\mathcal{M}_4$	1992	2119	160097
$\mathcal{M}_5$	4262	462715	531325
$\mathcal{M}_6$	3019	45559	281093
$\mathcal{M}_7$	3564	58259	231850

In contrast, the HESC method proposed on Advogato performs well. When the vital nodes are removed, the method has a strong destructive power to the network in Fig. 6(a). On Wiki-Vote, before removing Top50, the differences among the methods are small, and the node sorting performance of CC method is better. However, with the increase of the number of removed nodes, the advantages of HESC and HEC methods are shown, especially when the top 100 is



(a) DC-HESC (b) CC-HESC (c) BC-HESC

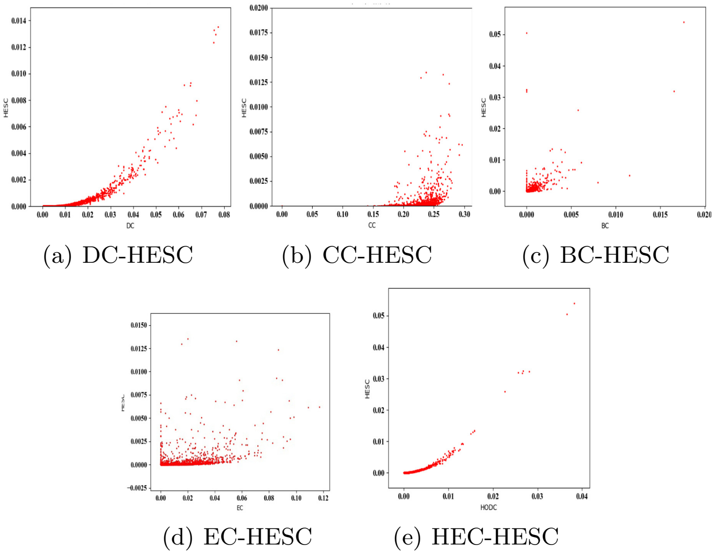
(d) EC-HESC (e) HEC-HESC

Fig. 3. The relationship between HESC and others in Advogato

removed, the relative size of strongly connected subgraphs in HESC method is the smallest in Fig. 6(b). The HEC and HESC methods proposed are close to each other on soc-Epinions1, and the performance is only better than CC method. There are some differences in the methods of identifying vital nodes based on high-order structure in Fig. 6. But as a whole, with the vital nodes removed, the more seriously the network is destroyed.

In order to better evaluate our proposed methods, we carries out experiments on SIR model to test the propagation ability of nodes, and compares it with other algorithms. Nodes in SIR epidemic transmission model have three possible states at any time: susceptible, infected and recovered. At the time t, the proportion of these three groups of people in the crowd is used $S(t)$, $I(t)$ and $R(t)$ to express separately. $S(t)$ represents the proportion of nodes in a network that are vulnerable to infection. $I(t)$ represents the ability to transmit disease to other vulnerable nodes in an infected state. Each infected node can randomly transmit disease to its neighbor nodes through a certain probability. $R(t)$ represents the proportion of nodes that have been infected but have recovered and have immunity. In the SIR model of complex networks, we assume that all neighbor nodes around infected nodes have the chance to be infected.

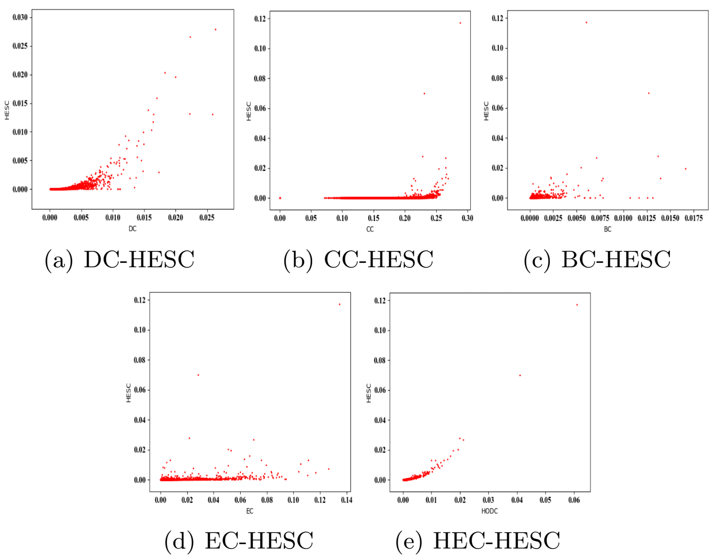
We used the Top-10, Top-50 and Top-100 nodes ranked by various centralities as infected nodes in the initial network. Then we used the proportion of infected nodes and recovered nodes in the network to judge the influence of nodes when the network reaches steady state and compare the differences between different methods. Figure 7, Fig. 8 and Fig. 9 shows the propagation ability of infected



(a) DC-HESC (b) CC-HESC (c) BC-HESC

(d) EC-HESC (e) HEC-HESC

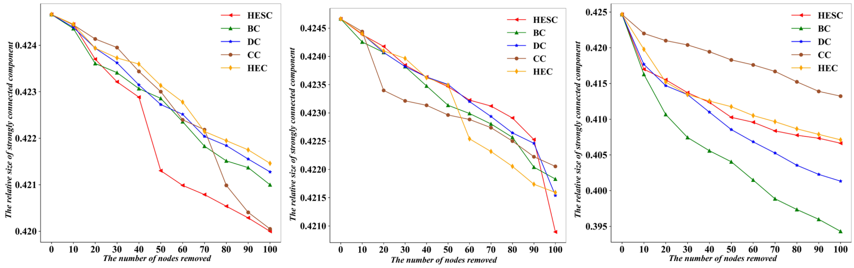
Fig. 4. The relationship between HESC and others in Wiki-Vote



(a) DC-HESC (b) CC-HESC (c) BC-HESC

(d) EC-HESC (e) HEC-HESC

Fig. 5. The relationship between HESC and others in soc-Epinions1



(a) Advogato (b) Wiki-Vote (c) soc-Epinions1

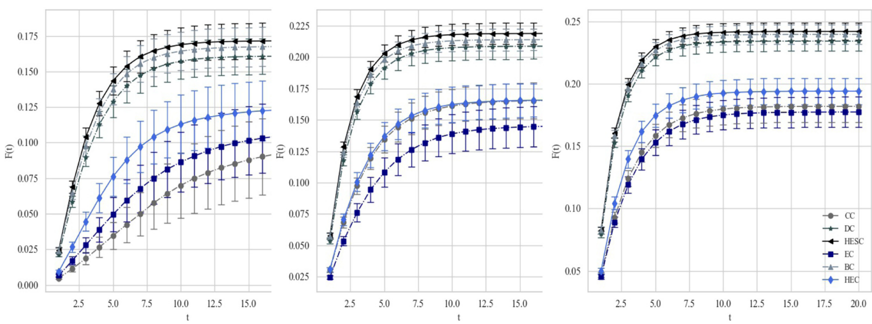
Fig. 6. The relative size of strongly connected component under static attack of different datum

nodes by the top-L nodes as ranked by six centrality measures under these three datasets.

Comparatively speaking, our proposed HESC is superior to the classical methods in both propagation range and propagation rate in Advogato dataset (see Fig. 7). HESC, BC and DC have almost the same performance on Advogato, because these two centralities are all positively correlated with HESC in this network (see Fig. 4). However, the propagation rate and range of HEC method are relatively poor, which may be due to the close number of various motifs in this data set. At this time, the advantage of HEC can not be reflected by taking the largest number of motifs in the network as the analysis object.

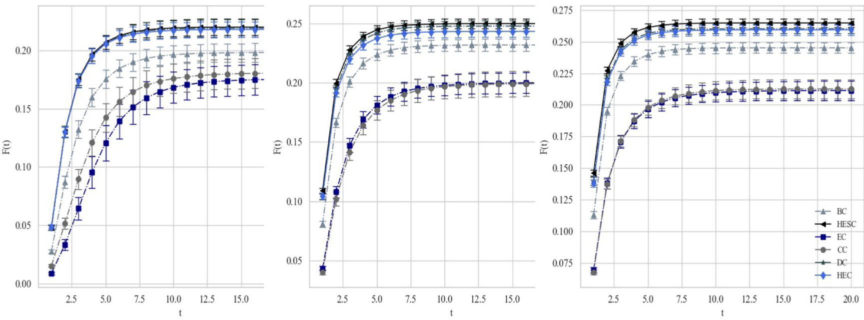
Figure 8(a) shows that HEC and HESC are basically similar to DC in terms of propagation range and propagation rate in Wiki-Vote. However, with the increase of infected nodes, HESC is better than traditional measurement methods in both transmission rate and transmission range (see Fig. 8(b) and 8(c)).

We can see that the transmission range and efficiency of HESC are better than other methods when Top10 node is used as a source of infection from Fig. 9.



(a) Top10 of Advogato (b) Top50 of Advogato (c) Top100 of Advogato

Fig. 7. Experiment of TOP Nodes as initial infectious source node in Advogato



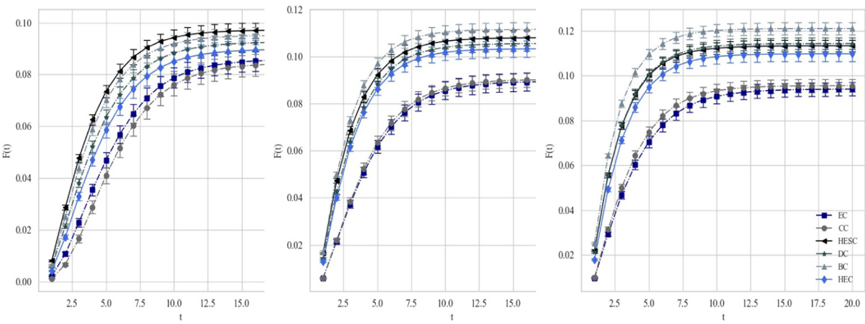
(a) Top10 of Wiki-Vote (b) Top50 of Wiki-Vote (c) Top100 of Wiki-Vote

Fig. 8. Experiment of TOP Nodes as initial infectious source node in Wiki-Vote

However, with the increase of infected nodes, the advantages of HESC gradually diminished. The propagation capability of Top nodes obtained by HEC is similar to that of HESC in Fig. 5. Furthermore, from the error bar of Fig. 9, we can see that the results are not sensitive to the dynamic process on networks.

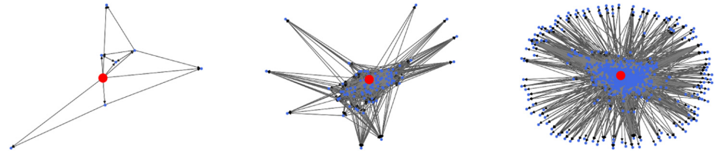
Although HESC method can select more important nodes by the analysis of the experimental results of different data sets, there are some differences for different networks. In soc-Epinions1 network, only Top10 nodes have better propagation ability than other sorting results, which may be due to the large proportion of strongly connected subgraphs in the network. Therefore, when the number of selected source nodes is large, BC's advantages appear.

Table 4, Table 5 and Table 6 show the TOP5 nodes obtained by different methods in three datasets. Ego network is composed of a centered ego, direct contacts namely alters, and the iterations among them. We select the most influential node of the three networks by different centrality measures from Fig. 10, Fig. 11 and Fig. 12 to get its Ego networks. We try to explain why HESC



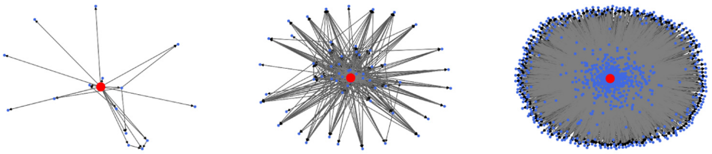
(a) Top10 of soc-Epinions1 (b) Top50 of soc-Epinions1 (c) Top100 of soc-Epinions1

Fig. 9. Experiment of TOP Nodes as initial infectious source node in soc-Epinions 1

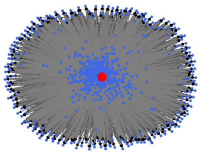


(a) v_{30} of Advogato (b) v_{46} (c) v_{157}

Fig. 10. Ego-network of Top1 node by various methods in Advogato

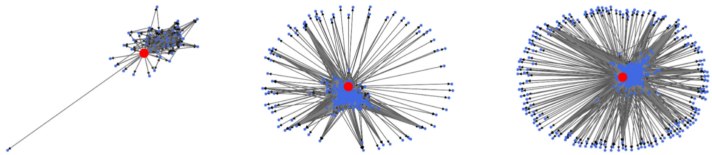


(a) v_{4037} (b) v_{2398} (c) v_{2565}



(d) v_{766}

Fig. 11. Ego-network of Top1 node by various methods in Wiki-Vote



(a) v_{18} (b) v_{44} (c) v_{645}

Fig. 12. Ego-network of Top1 node by various methods in soc-Epinions1

Table 4. Top-5 nodes by different methods in Wiki-Vote

TOP	DC	BC	CC	EC	HEC	HESC
TOP1	2565	4037	2565	2398	2565	766
TOP2	1549	15	1549	4037	766	2565
TOP3	766	2398	15	15	11	1549
TOP4	11	1549	72	4191	2688	457
TOP5	1166	2535	737	2625	1549	11

Table 5. Top-5 nodes by different methods in Advogato

TOP	DC	BC	CC	EC	HEC	HESC
TOP1	157	46	157	46	30	157
TOP2	46	30	46	30	328	597
TOP3	597	328	597	328	126	232
TOP4	30	286	172	438	286	593
TOP5	328	719	328	719	172	793

Table 6. Top-5 nodes by different methods in soc-Epinions1

TOP	DC	BC	CC	EC	HEC	HESC
TOP1	18	18	44	18	18	645
TOP2	645	737	763	401	645	634
TOP3	634	136	634	550	634	44
TOP4	763	790	2066	737	143	71399
TOP5	143	143	645	34	790	763

outperforms others by Ego network in these three networks intuitively. The large solid circle in the center of Fig. 10, Fig. 11 and Fig. 12 is the most influential node obtained by various measurements. Clearly, the Ego network of the most influential node in HESC is more compact than other methods. To some extent, this reflects that when TOP nodes as initial infectious source nodes why the spreading rate of HESC is faster than that of others, and the total number of infected nodes of HESC is also larger than that of others.

5 Conclusions

This paper reconstructs the initial network by high-order structure. In order to describe the high-order information such as the interaction between nodes in the network, it defined the concept of high-order. By fusing the low-order information and high-order information of nodes, the HESC was proposed to identify the vital nodes. At the same time, the network topology and the propagation dynamic model was used to evaluate the node ranked. The results of experiments demonstrate that even though not all the nodes in initial network can form a high-order structure, the nodes that form a higher-order structure play a more influential role in the network. The influential nodes based on the high-order structure show slightly different performance in different networks. The propagation ability are better than that compared with influential nodes identified by other algorithms.

References

1. Zhang, J., Tang, J.: Survey of social influence analysis and modeling. Sci. Sin. Inform. **47**, 967–979 (2017) (in Chinese). https://doi.org/10.1360/N112017-00137
2. Han, Z.M., Chen, Y., Liu, W., Yuan, B.H., Li, M.Q., Duan, D.G.: Research on node influence analysis in social networks. Ruan Jian Xue Bao/J. Softw. **28**(1), 84–104 (2017). (in Chinese). http://www.jos.org.cn/1000-9825/5115.htm
3. Wasserman, S.: Social network analysis methods and applications. Contemp. Sociol. **91**(435), 219–220 (1995)
4. Bonacich, P.F.: Factoring and weighting approaches to status scores and clique identification. J. Math. Sociol. **2**(1), 113–120 (1972)
5. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**(3), 215–239 (1978)
6. Freeman, L.C.: A set of measures of centrality based on betweenness. Sociometry **40**(1), 35–41 (1977)
7. Gao, C., Wei, D., Hu, Y., et al.: A modified evidential methodology of identifying influential nodes in weighted networks. Phys. A **392**(21), 5490–5500 (2013)
8. Chen, D., Lü, L., Shang, M., et al.: Identifying influential nodes in complex networks. Phys. A **391**(4), 1777–1787 (2012)
9. Wei, D., Deng, X., Zhang, X., et al.: Identifying influential nodes in weighted networks based on evidence theory. Phys. A **392**(10), 2564–2575 (2013)
10. Benson, A.R., Gleich, D.F., Leskovec, J.: Higher-order organization of complex networks. Science **353**(6295), 163–166 (2016)
11. Benson, A.R.: Tools for higher-order network analysis (2018)
12. Milo, R.: Network motifs: simple building blocks of complex networks. Science **298**(5594), 824–827 (2002)
13. Shen-Orr, S.S., Milo, R., Mangan, S., et al.: Network motifs in the transcriptional regulation network of Escherichia coli. Nat. Genet. **31**(1), 64–68 (2002)
14. Yin, H., Benson, A.R., Leskovec, J., et al.: Local higher-order graph clustering. In: KDD 2017, pp. 555–564 (2017)
15. Zhao, H., Xiaogang, X., Yangqiu, S., et al.: Ranking users in social networks with higher-order structures. In: AAAI (2018)
16. Shafer, G.: A Mathematical Theory of Evidence. by Glenn Shafer, vol. 73, no. 363, pp. 677–678 (1978)
17. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**(2), 325–339 (1967)
18. Schank, T., Wagner, D.: Finding, counting and listing all triangles in large graphs, an experimental study, vol. 3503, pp. 606–609. Springer, Heidelberg (2005)
19. Ren, X.L., Lü, L.Y.: Review of ranking nodes in complex networks. Chin. Sci. Bull. (Chin. Ver.) **59**(1197), 1175–1280 (2014) (in Chinese). https://doi.org/10.1360/72013-1280
20. Yan, G., Zhang, M., Luo, H., et al.: Identifying vital nodes algorithm in social networks fusing higher-order information. J. Commun. **40**(10), 109–118 (2019). (in Chinese)
21. Bian, T., Deng, Y.: A new evidential methodology of identifying influential nodes in complex networks. Chaos Solitons Fract. **103**, 101–110 (2017)
22. Kunegis, J.: KONECT - the koblenz network collection. In: Proceedings of the International Conference on World Wide Web Companion, pp. 1343–1350 (2013)
23. Leskovec, J., Huttenlocher, D., Kleinberg, J.: Signed networks in social media. In: CHI 2010 (2010)
24. Richardson, M., Agrawal, R., Domingos, P.: Trust management for the semantic web. In: ISWC (2003)