Information Processing and Management 000 (2016) 1–12

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Efficient identification of node importance in social networks

Yunyun Yang, Gang Xie*

College of Information Engineering, Taiyuan University of Technology, Taiyuan, Shanxi 030024, People's Republic of China

ARTICLE INFO

Article history:
Received 6 December 2015
Revised 22 March 2016
Accepted 3 April 2016
Available online xxx

Keywords:
Social networks
Node importance
Comprehensive decision analysis method

ABSTRACT

In social networks, identifying influential nodes is essential to control the social networks. Identifying influential nodes has been among one of the most intensively studies of analyzing the structure of networks. There are a multitude of evaluation indicators of node importance in social networks, such as degree, betweenness and cumulative nomination and so on. But most of the indicators only reveal one characteristic of the node. In fact, in social networks, node importance is not affected by a single factor, but is affected by a number of factors. Therefore, the paper puts forward a relatively comprehensive and effective method of evaluation node importance in social networks by using the multi-objective decision method. Firstly, we select several different representative indicators given a certain weight. We regard each node as a solution and different indicators of each node as the solution properties. Then through calculating the closeness degree of each node to the ideal solution, we obtain evaluation indicator of node importance in social networks. Finally, we verify the effectiveness of the proposed method experimentally on a few actual social networks.

1. Introduction

In the era of big data, social networks can be found anywhere. Social networks can be seen as a set of the nodes joined in pairs by edges. In recent years, the analytical approach of the social networks has attracted considerable curiosity from the social and the scientific community. This curiosity can be attributed to a fascinating focus, namely the relationship between social entities, as well as the pattern and meaning of the relationship (Chen & Wang, 2012; Kossinets & Watts, 2006; Van Noorden, 2014; Wasserman, 1994; Watts & Strogatz, 1998). In particular, the degree of importance of each node in social networks is not the same, that is, different nodes have different weights. Therefore, in the actual network, it is valuable to find paramount nodes for controlling the entire network. Consequently, we need to capture and reveal the relationship between the feature of the nodes and network topology so that we understand the impact of a node on system functions, such as reliability, robustness and controllability (Ahlswede, Cai, & Li, 2000; de Camargo, Neto, & Pires, 2012; Hawick & James, 2007).

In social networks, there are a variety of measurement indicators for identifying the most crucial nodes (roles). In general, the measurement indicators (Arulselvan, Commander, & Elefteriadou, 2009; Bhola, Grover, & Sinha, 2010; Borgatti, 2006; Hewett, 2011; Hu, Wang, & Lee, 2010; Xie, Li, & Zhang, 2015) are divided into two categories: from the perspective of local attributes and the perspective of global attributes. For example, degree centrality (Bonacich, 1972) represents that the more the number of neighboring nodes are, the more important the nodes are, which is the simplest index of centrality. Information index (Bonacich, 1987) depends on the amount of information through its propagation path. Besides,

* Corresponding author. Tel.: +86 03516018860, fax: +8618734176603.
E-mail address: xiegang189@163.com (G. Xie).

http://dx.doi.org/10.1016/j.ipm.2016.04.001
0306-4573/© 2016 Published by Elsevier Ltd.

Kitsak, Gallos, and Havlin (2010) proposed k-shell decomposition. Closeness centrality (Freeman, 1979) discussed by Rothenberg et al. is a global measure of centrality based on the distance between nodes in a connected network. Closeness centrality can be seen as a measure of how long it will take information to spread from a given node to others in the network. In addition, there are a host of other concepts such as subgraph centrality (Estrada & Rodriguez-Velazquez, 2005), eigenvector centrality Stephenson and Zelen (1989) and cumulative nomination (Poulin, Boily, & Masse, 2000). In those regards, Ren and Lv (2014), Wang and Zhang (2006), and He, Li, Gan, and Zhu (2008) have done excellent summaries. Further, Sun and Luo (2012), Liu, Ren, Guo, and Wang (2013) also summarized the history of the methods of mining important nodes in complex networks and summarized the research results. They are trying to describe and measure the characteristics of a node from a different point of view, and adapt for different networks (Liu, Tang, & Zhou, 2015; London, Németh, & Pluhár, 2015; Ren, Liu, & Shao, 2013; Zhang, Xu, & Li, 2011). That is, they only reflect one perspective of node importance. But in modern society networks, node importance is not only determined by one attribute, but is also determined by a number of attributes. As an example, individuals' I.Q is determined by education, age, experience, etc. In addition, the proportion of each attribute is not equal. We should give different weights for each attribute. The literature (Yu, Liu, & Li, 2013) combined some indicators to discuss node importance and mainly considered structural characteristics of the network topology. In social networks, node importance should combine with some information concerning the node itself in the actual networks.

Based on this, in this paper from the network topology and the center of one node in social networks, we consider five indicators, including prestige, the amount of information dissemination, the transmitting path of the information, and hence propose a comprehensive evaluation of node importance for social networks.

The rest of the paper is built up as follows. Section 2 introduces the notations and notions related to our problem. Section 3 presents a scheme for efficient identification of node importance in social networks. In Section 4, this developed scheme is applied to identify the most important node(s) in some networks. In Section 5 the concluding remarks are given.

2. The definition of node importance indicators

According to the emphasises of different indicators, from more than one perspective (including the connection status between the nodes, the path of the target node to other nodes, and the perspective of simulation of the flow analysis), we choose five indicators: structural holes, flow betweenness centrality, cumulative nomination, information indicator, and subgraph centrality, to evaluate node importance comprehensively. The paper essentially studies a multi-attribute ranking problem. The definitions of these indicators are described in this section.

First of all, a short review of the required basic graph terminology is presented. For the sake of simplicity, we shall limit ourselves to undirected and unweighted networks. In this situation, if a node i is linked to another node j, then node j is necessarily linked to node i. Besides, all the links have the same weight.

Let G be a graph. A graph G consists of a non-empty finite set $V(G)$ of elements called nodes, and a finite set $E(G)$ of ordered pairs of different nodes called edges. We write $G = (V, E)$ where V and E are the node set and edge set of G, respectively. The size of G is the number of nodes in G, and is denoted by $|G|$. $V = \{v_1, v_2, \cdots, v_N\}$, $|V| = N$; $E = \{e_1, e_2, \cdots, e_M\} \subseteq V \times V$, $|E| = M$.

Next, present some measures typically used to identify node importance of social networks.

Definition 1 (Structural holes (Burt, 2009))**.** In networks, when there is no direct connection between two nodes, in the same vein, and there is no redundancy relationship between them, the obstacle between them is a structural hole. Burt (2009) gave the expression of structural holes with network constraints coefficient.

$$CH(i) = \sum_{j} (P_{ij} + \sum_{q \neq i \neq j} P_{iq} P_{qj})^2. \tag{1}$$

Where q is indirect node connecting node i and node j; P_{ij} means the proportion of node i to spend time (energy) on the node j and its total time (energy). The lower $CH(i)$ is, the greater the degree of structural holes is. That is to say the location of the node i is more important.

Definition 2 (Flow betweenness centrality (Freeman, Borgatti, & White, 1991))**.** In most networks, however, information (or anything else) does not flow only along geodesic paths. News or a rumor or a message does not know the ideal route to take from one place to another more randomly. To work around these issues, in 1991 Freeman et al. raised flow betweenness. Flow betweenness is based on the idea of maximum flow. Imagine each edge in a network as a pipe that can carry a unit flow of some fluid. We can ask what the maximum possible flow then is between a given source node s and target node t through these pipes. In general, the answer is that more than a single unit of flow can be carried between the source and target by making simultaneous use of several different paths through the network. The flow betweenness of a node i is defined as the amount of flow through node i when the maximum flow is transmitted from node s to node t. It is a measure of the influence a node has over the spread of information through the network.

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

where $g_{jk}(i)$ is the number of geodesic paths from node j to node k that pass through node i, and g_{jk} is the total number of geodesic paths from node j to node k.

Definition 3 (Cumulative nomination (Poulin et al., 2000))**.** In the initial stage, each node is assigned an initial nomination. Assume that more central individuals will be nominated more often. In every subsequent iteration, new nomination of the node includes its previously itself nomination and the other nodes' nomination connected to the node. After several iterations, the nomination of the node will be closeness to a constant, that is the final nomination of the nodes. Cumulative nomination refers to the reputation of a node itself in the network. Thus cumulative nomination can be used to measure node importance. Firstly, suppose there are n nodes in the network. The calculation process of cumulative nomination is as follows:

$$q_i^{n+1} = \frac{q_i^n + \sum_j a_{ij} q_j^n}{\sum_k [q_k^n + \sum_j a_{kj} q_j^n]}, \tag{3}$$

where

$$\sum_i q_i^n = 1, \quad q_i^0 = 1.$$

Mathematically, cumulative nomination of the node i can be written as the following expression:

$$CN(i) = N \lim_{n\to\infty} q_i^n. \tag{4}$$

Cumulative nomination summarizes the contact structure of social networks based on the iteration of a mapping.

Definition 4 (Information indices (Guo & Lu, 2012))**.** Let $\boldsymbol{M}$ be an adjacency matrix describing the connected network, $\boldsymbol{D}$ a diagonal matrix of the degree of each point and $\boldsymbol{J}$ a matrix with all its elements equal to 1. The indicator of the centrality of information is calculated by inverting the matrix $\boldsymbol{B}$ defined by:

$$\boldsymbol{C} = (c_{ij})_{N\times N} = \boldsymbol{B}^{-1} = (\boldsymbol{D} - \boldsymbol{M} + \boldsymbol{J})^{-1}.$$

The information matrix is given explicitly by:

$$q_{ij} = (c_{ii} + c_{jj} - 2c_{ij})^{-1}, \tag{5}$$

where the values q_{ij} summarize the information contained in all feasible paths between node i and node j. Consequently the centrality indicator of information indices can be formalized as

$$CI(i) = \frac{1}{N} \sum_{j\neq i} q_{ij}. \tag{6}$$

Information indicator, which can be used to characterise node i within a connected network or within a connected component of a network, is a measure of node importance through the spread of the path information.

Definition 5 (Subgraph centrality (Estrada & Rodriguez-Velazquez, 2005))**.** On the basis of the number of closed walks starting and ending at the node, the "subgraph centrality" $CS(i)$ for nodes in a network has been proposed. Closed walks are appropriately weighted so that their influence on the centrality decreases as the order of the walk increases. Each closed walk is associated with a connected subgraph, which means that this measure counts the times that a node takes part in the different connected subgraphs of the network, with smaller subgraphs having higher importance. Consequently, the measure of the "subgraph centrality" $CS(i)$ for nodes in a network is given by the expression

$$CS(i) = \sum_{t=1}^{\infty} \frac{a_{ii}^t}{t!} = \sum_{j=1}^{N} (\xi_j^i)^2 e^{\lambda_j}, \tag{7}$$

where $\lambda_j(1, 2, \cdots, N)$ is the eigenvalues of the adjacency matrix $\boldsymbol{A}$, ξ_j means the eigenvector corresponding to the eigenvalue λ_j, and ξ_j^i means the ith element of the eigenvector ξ_j. It manifests that the subgraph centrality $CS(i)$ can be obtained mathematically from the spectra of the adjacency matrix of the network. Compared with other indicators, the subgraph centricity is more accurate for differentiating node importance.

The five indicators mentioned above, structural holes and subgraph centrality are considered from the structure of the network; flow betweenness centrality is mainly considered from the perspective of simulation of the flow analysis; information index is focused on the simulation of the flow analysis; cumulative nomination stresses on the hierarchies and social in the networks. From different points of view, these five indicators consider node importance, respectively. In this paper, we give different weights to them, and then integrate them into a comprehensive indicator to measure the node importance. The details will be given in the next section.

3. Comprehensive analysis of node importance

Combining the thought and theory of multi-objective decision method, we put forward a relatively comprehensive and effective method of evaluating node importance in social networks. In this method, we regarded each node as a solution, and regard multiple indicators of the node as the solution properties. By calculating the closeness degree between each node and the ideal solution, we obtained evaluation indicators of key nodes in complex networks. Firstly, we give a generic description and contextualization of the proposed method.

The essence of the comprehensive decision analysis method is selecting the best alternative from a series of alternatives with multiple attributes to meet certain goals. The process of the comprehensive decision analysis method involves primarily the following aspects: the normalization decision matrix, the determination of the weight of each attribute and the sorting of the comprehensive scores of the evaluated schemes.

The decision matrix must be normalized because these decision attributes are contradictory, and each attribute dimension and order of magnitude are usually different. In order to avoid such conflicts, it is necessary to make a normalization for the decision matrix. Then, according to certain rules, give a certain weight to each attribute. At last, by using multiple criteria method, we calculate the comprehensive scores of the evaluated schemes and sort the scheme according to their scores.

3.1. Constitution of the normalized decision matrix

Suppose that there are n nodes in social networks. The set of decision-making plan can be written as $\boldsymbol{A} = \{\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_N\}$. Assume that there are m evaluation importance indicators for each node. Accordingly, the attribute set of decision-making plan is symbolized by $Q = \{Q_1, Q_2, \cdots, Q_m\}$. $A_i(Q_j)$ stands for the value of the jth indicator for the ith node, where $i = 1, 2, \cdots, N$, and $j = 1, 2, \cdots, m$. The decision matrix can be obtained as follows:

$$\boldsymbol{X} = \begin{pmatrix} A_1(Q_1) & \cdots & A_1(Q_m) \\ \vdots & \ddots & \vdots \\ A_N(Q_1) & \cdots & A_N(Q_m) \end{pmatrix}. \tag{8}$$

A scheme contains a number of indicators, divided into cost indices (the higher the index, the less the efficiency) and benefit indices (the higher the index, the better the efficiency). Moreover, dimension is different for each indicator. Among the above mentioned five indicators, structural holes belong to cost index, and the remaining four indicators belong to benefit index. In order to compare, we make the following process for the decision matrix.

For cost indices,

$$r_{ij} = A_i(Q_j)^{\min} / A_i(Q_j). \tag{9a}$$

Similarly, for efficiency indices, one has

$$r_{ij} = A_i(Q_j) / A_i(Q_j)^{\max}. \tag{9b}$$

Here

$$A_i(Q_j)^{\min} = \min\{A_i(Q_j) | 1 \le i \le N\},$$

$$A_i(Q_j)^{\max} = \max\{A_i(Q_j) | 1 \le i \le N\}.$$

Let $\boldsymbol{R} = (r_{ij})_{N\times m}$ be the normalized decision matrix, and $w_j (j = 1, 2, \cdots, m, \Sigma w_j = 1)$ is the weight of the jth indicator. Hence, normalized matrix $\boldsymbol{Y}$ with weights can be written as

$$\boldsymbol{Y} = (y_{ij})_{N\times m} = \begin{pmatrix} w_1 r_{11} & \cdots & w_m r_{1m} \\ \vdots & \ddots & \vdots \\ w_1 r_{N1} & \cdots & w_m r_{Nm} \end{pmatrix}. \tag{10}$$

Firstly, compare every two indicators through three scale method (0, 1, 2), we can get the comparison matrix $\boldsymbol{B}$. In the network, *CH* does not only reflect the global features of the network, but also distinguishes the local differences of nodes extremely well, and thus *CH* is set to be the most important indicator. The greater the prestige of a node, the greater its influence, so *CI* is the second important. It is arduous to compare the merits of *CF* and *CN*, because the former incarnates the amount of information of the node in the network, while the latter gives expression to the center feature of the node. Hence, this paper makes *CF* and *CN* equally important. While compared with other indicators, *CS* reflects little information regarding the network structure. Accordingly *CS* is the least important.

Comprehensive analysis can come to the conclusion: $CH > CI > CF = CN > CS$, as indicated in Table 1

In Table 1, $\boldsymbol{B} = (b_{ij})_{5\times 5}$. When indicator j is more important than indicator i, $b_{ij} = 0$. If indicator i and indicator j are equally important, then $b_{ij} = 1$. Otherwise, $b_{ij} = 2$.

Then, we refer to Analytic Hierarchy Process (Zhu, Meng, & Kan, 1999) to calculate the weight of each indicator, and the corresponding computer program code is in the following.

Table 1
The comparative results of the indicators.

B	*CH*	*CF*	*CN*	*CI*	*CS*
CH	1	2	2	2	2
CF	0	1	1	0	2
CN	0	1	1	0	2
CI	0	2	2	1	2
CS	0	0	0	0	1

```
B=[1 2 2 2 2; 0 1 1 0 2; 0 1 1 0 2; 0 2 2 1 2; 0 0 0 0 1]
for i=1:size(B,1)
    r(i)=sum(B(i,:));
end
R=max(r)-min(r);
for i=1:size(B,1)
   for j=1:size(B,2)
       c(i,j)=9^((r(i)-r(j))/R);
    end
end
disp(c);
M=prod(c,2);
W = M.^(1/5);
w = W./sum(W(:,:))
```

In accordance with the range method, construct the judgment matrix of comparison matrix ***B***. Further through the consistency check, get the weight of each index respectively as follows: $w_{CH} = 0.4556, w_{CI} = 0.2630, w_{CF} = 0.1154, w_{CN} = 0.1154, w_{CS} = 0.0506$.

3.2. Calculation of the closeness degree to the ideal solution

The negative-ideal solution, denoted as A^-, and the positive-ideal solution, denoted as A^+, are defined based on the matrix ***Y*** as follows:

$$A^- = \{y_{i1}^{\min}, y_{i2}^{\min}, \cdots, y_{im}^{\min}\}, \tag{11a}$$

$$A^+ = \{y_{i1}^{\max}, y_{i2}^{\max}, \cdots, y_{im}^{\max}\}. \tag{11b}$$

Thus, in the light of Euclidean distance, calculate the distances D_i^- and D_i^+ of each solution A_i to the negative-ideal solution A^- and the positive-ideal solution A^+ by Eqs. (12a) and (12b), respectively.

$$D_i^- = [\sum_{j=1}^{m}(y_{ij} - y_j^{\min})^2]^{1/2}, \tag{12a}$$

$$D_i^+ = [\sum_{j=1}^{m}(y_{ij} - y_j^{\max})^2]^{1/2}. \tag{12b}$$

By Eq. (13), the closeness degree Z_i to the ideal solution is calculated. Then rank the alternatives in accordance with the closeness degree to the ideal solution. The higher Z_i are assumed to be more important and should be given higher priority.

$$Z_i = D_i^- / (D_i^- + D_i^+), \ 0 \le Z_i \le 1. \tag{13}$$

3.3. The specific steps to evaluate node importance in actual social networks

Input The comprehensive decision matrix $\boldsymbol{X} = (x_{ij})_{N\times m} = (CH_i, CF_i, CN_i, CI_i, CS_i)_{N\times m}$.
Output The value Z_i of the importance of node i.
Step 1 By Eq. (9) calculate the standard matrix $\boldsymbol{R} = (r_{ij})_{N\times m}$;
Step 2 Calculate the weight of each indicator. After substituting Eq. (10), construct weighted normalized matrix ***Y***;
Step 3 From Eqs. (11a) and (11b), obtain the negative-ideal solution A^- and positive-ideal solution A^+, respectively;

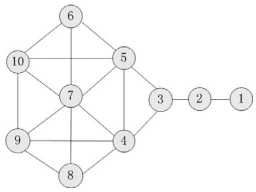

Fig. 1. Kite network.

Table 2
Calculation results of various indicators in "Kite network".

ID	*CH*	*CF*	*CN*	*CI*	*CS*
1	1.25	0.4011	0.0398	0.0000	0.5942
2	0.5556	0.5926	0.1712	8.0000	1.3542
3	0.4944	1.0624	0.6976	19.7973	4.2341
4	0.4701	1.6438	1.4165	27.2736	13.1646
5	0.4701	1.6438	1.4165	27.2736	13.1646
6	0.7059	1.2885	1.0181	16.5659	7.1571
7	0.4746	1.8039	1.7133	28.7362	18.6568
8	0.7059	1.2885	1.0181	16.5659	7.1571
9	0.5783	1.5043	1.2545	23.3124	10.4007
10	0.5783	1.5043	1.2545	23.3124	10.4007

Step 4 In accordance with Eqs. (12a) and (12b), calculate the distances D_i^- and D_i^+ of each solution A_i to the negative-ideal solution A^- and the positive-ideal solution A^+, respectively;
Step 5 On the basis of Eq. (13), calculate the closeness degree Z_i to the ideal solution, namely, the comprehensive importance of a node;
Step 6 Rank the alternatives from big to small according to the closeness degree Z_i to the ideal solution. The higher the value of Z_i, the greater the position of the node in the network.

The time complexity of the process is mainly decided by two aspects. On the one hand, the time complexity of the process depends on the complexity of the five selected indicators. On the other hand, the time complexity of the process is determined by the time complexity of the whole sorting scheme.

The time complexity of the process mainly depends on the complexity of the five selected indicators which are *CH, CF, CN, CI* and *CS*. The time complexities of the five selected indicators are given as $O(N)$, $O(N^2)$, $O(N+M)$, $O(N^3)$ and $O(N^2)$, respectively. And the time complexity of the whole sorting scheme is $O(Nlog2N)$. Obviously, the time complexity of the proposed methodology is not high. So it could be applied in large-scale networks.

4. Example explanation

The effectiveness of the method is demonstrated by the following examples.

4.1. Example 1

Take "Kite network" (Wang, Li, & Chen, 2012) as an example (as shown in Fig. 1), calculate each indicator of all nodes, as indicated in Tab. 2.

In Table 2, the attribute values of all of the indicators worked out. Further, the weighted normalized matrix can be got by using the weight of each indicator and the normalized decision matrix $\boldsymbol{Y}$:

$$\boldsymbol{Y} = \begin{pmatrix} 0.1713 & 0.0257 & 0.0027 & 0 & 0 \\ 0.3855 & 0.0379 & 0.0115 & 0.0732 & 0.0217 \\ 0.4332 & 0.0680 & 0.0470 & 0.1812 & 0.0537 \\ 0.4556 & 0.1052 & 0.0954 & 0.2496 & 0.0740 \\ 0.4556 & 0.1052 & 0.0954 & 0.2496 & 0.0740 \\ 0.3034 & 0.0824 & 0.0686 & 0.1516 & 0.0449 \\ 0.4513 & 0.1154 & 0.1154 & 0.2630 & 0.0779 \\ 0.3034 & 0.0824 & 0.0686 & 0.1516 & 0.0449 \\ 0.3704 & 0.0962 & 0.0845 & 0.2134 & 0.0632 \\ 0.3704 & 0.0962 & 0.0845 & 0.2134 & 0.0632 \end{pmatrix},$$

Thus, the negative-ideal decision scheme A^- is

$$\{0.1713 \quad 0.0257 \quad 0.0027 \quad 0 \quad 0\};$$

Table 3
The comprehensive decision results of node importances for "Kite network".

ID	D_i^-	D_i^+	Z_i
1	0	0.4205	0
2	0.2279	0.2468	0.4801
3	0.3287	0.1213	0.7304
4	0.4044	0.0264	0.9386
5	0.4044	0.0264	0.9386
6	0.2236	0.1998	0.5281
7	0.4176	0.0043	0.9898
8	0.2236	0.1998	0.5281
9	0.3175	0.1062	0.7494
10	0.3175	0.1062	0.7494

Table 4
Calculation results of some indicators in "Kite network" and the sorting of nodes.

ID	Betweenness	Degree	Closeness	$Sort_B$	$Sort_D$	$Sort_C$	$Sort_{Theproposedmethod}$
1	0.000	0.111	0.034	6	6	6	7
2	0.356	0.222	0.048	3	5	5	6
3	0.622	0.333	0.067	1	4	2	4
4	0.370	0.556	0.071	2	2	1	2
5	0.370	0.556	0.071	2	2	1	2
6	0.000	0.333	0.056	6	4	4	5
7	0.163	0.667	0.067	4	1	2	1
8	0.000	0.333	0.056	6	4	4	5
9	0.037	0.444	0.059	5	3	3	3
10	0.037	0.444	0.059	5	3	3	3

and the positive-ideal decision scheme A^+ is

$$\{0.4556 \quad 0.1154 \quad 0.1154 \quad 0.2630 \quad 0.0779\}.$$

Finally, from Eqs. (12a), (12b) and (13), one has D_i^-, D_i^+ and Z_i.

The sort of the closeness degree Z_i can be concluded as in Tab. 3, that is $Z_7 > Z_4 = Z_5 > Z_9 = Z_{10} > Z_3 > Z_6 = Z_8 > Z_2 > Z_1$.

From Fig. 1, node 7 is in the position of the largest network information traffic. Node 7 can timely and accurately capture and transfer of information. Node 4 and node 5 are in the same location in the network structure, so they have the same sort. Moreover, node 4 and node 5 are located at the confluence of the center and the secondary node, which play the roles of the bridge. Therefore, node 4 and node 5 rank the second. Compared with node 4 and node 5, node 9 and node 10 do not play the role of a bridge. Therefore, node 9 and node 10 rank behind node 4 (node 5). Although node 3 has low degree, if we delete node 3, the network can be in interruption completely. Thus, node 3 is more important than node 6 (or node 8), and node 2 (or node 1). For node 6 and node 8, they are located in the same location in the network structure. In the same vein, after deleting them, the connectivity of the network would be unchanged, so node 6 and node 8 rank more behind relatively. Intuitively, for node 2 and node 1, they make sense that less information gets transmitted, so they are less important.

Table 4 shows the comparative results of node importance indicators and their priorities by using a single indicator, such as betweenness, degree, and closeness. The sorted result by using the one single indicator was different from the sorted result by using the proposed method. Because each indicator has a different emphasis. Betweenness is a global characteristic and reflects the influence of a node in the entire network. A node, which has a great betweenness value, is similar to bridge node and may be the junction of two communities. Degree is the local influence of a node in a network. The degree of a node represents the communication capability of a node with other nodes. The closeness centrality of a node can be expressed as the reciprocal of the total length of the shortest paths from the node to all the other nodes in the network. If closeness centrality of a node is large, the node lies to the network center, which indicates that a node occupies an important position in the network.

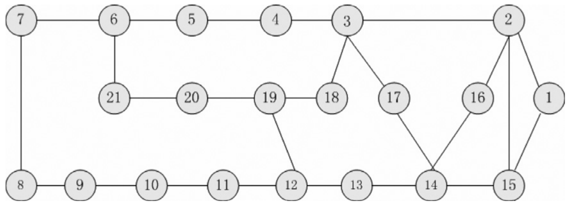

Fig. 2. The network topology of ARPA.

Table 5
The comprehensive decision results of node importances for ARPA.

ID	D_i^-	D_i^+	Z_i	ID	D_i^-	D_i^+	Z_i
1	0.3204	0.0723	0.1842	12	0.1476	0.2042	0.5803
2	0.1329	0.2248	0.6284	13	0.2473	0.0986	0.285
3	0.0273	0.3308	0.9237	14	0.0377	0.303	0.8893
4	0.2553	0.0859	0.2517	15	0.225	0.1378	0.3798
5	0.2648	0.0773	0.226	16	0.2605	0.0981	0.2735
6	0.1629	0.1972	0.5477	17	0.2511	0.1007	0.2862
7	0.2688	0.0769	0.2224	18	0.2533	0.0885	0.2588
8	0.2716	0.0743	0.2147	19	0.1507	0.1989	0.5689
9	0.2714	0.0743	0.215	20	0.2665	0.0769	0.2239
10	0.2701	0.0748	0.217	21	0.2696	0.0758	0.2195
11	0.265	0.0784	0.2284				

4.2. Example 2

Fig. 2 is the topology structure of ARPA (Advanced Research Project Agency). It is often used to analyze node importance. Here, in the light of the algorithm steps, we get the weighted normalized matrix $\boldsymbol{Y}$:

$$\boldsymbol{Y} = \begin{pmatrix} 0.1573 & 0.0303 & 0.0740 & 0.1945 & 0.0143 \\ 0.3255 & 0.0881 & 0.1154 & 0.2630 & 0.0193 \\ 0.4556 & 0.1154 & 0.0881 & 0.2610 & 0.0191 \\ 0.2307 & 0.0586 & 0.0359 & 0.1922 & 0.0141 \\ 0.2275 & 0.0586 & 0.0156 & 0.1932 & 0.0142 \\ 0.3374 & 0.0928 & 0.0094 & 0.2349 & 0.0172 \\ 0.2275 & 0.0604 & 0.0040 & 0.1939 & 0.0142 \\ 0.2250 & 0.0591 & 0.0022 & 0.1952 & 0.0143 \\ 0.2250 & 0.0590 & 0.0024 & 0.1960 & 0.0144 \\ 0.2250 & 0.0601 & 0.0048 & 0.1963 & 0.0144 \\ 0.2275 & 0.0624 & 0.0115 & 0.1956 & 0.0143 \\ 0.3399 & 0.0999 & 0.0286 & 0.2378 & 0.0174 \\ 0.2334 & 0.0773 & 0.0431 & 0.1917 & 0.0141 \\ 0.4240 & 0.1078 & 0.0963 & 0.2613 & 0.0192 \\ 0.2410 & 0.0609 & 0.0987 & 0.2267 & 0.0166 \\ 0.2250 & 0.0327 & 0.0732 & 0.1854 & 0.0136 \\ 0.2368 & 0.0353 & 0.0637 & 0.1853 & 0.0136 \\ 0.2334 & 0.0542 & 0.0401 & 0.1897 & 0.0139 \\ 0.3399 & 0.0876 & 0.0280 & 0.2327 & 0.0171 \\ 0.2275 & 0.0594 & 0.0122 & 0.1910 & 0.0140 \\ 0.2275 & 0.0584 & 0.0075 & 0.1875 & 0.0137 \end{pmatrix},$$

Table 5 gives the sort of the closeness degree Z_i for Fig. 2, that is $Z_3 > Z_{14} > Z_2 > Z_{12} > Z_{19} > Z_6 > Z_{15} > Z_{17} > Z_{13} > Z_{16} > Z_{18} > Z_4 > Z_{11} > Z_5 > Z_{20} > Z_7 > Z_{21} > Z_{10} > Z_9 > Z_8 > Z_1$.

In Tab. 6, we give the rankings of Top5% nodes by the proposed method and other methods mentioned in the literatures (Zhao, Wang, Zheng, & Guo, 2009; Zhou, Zhang, & Li, 2012) in ARPA. Figs. 3, 4 and 5 are part of ARPA after removing the Top5% nodes and their connected edges of ARPA. As can be observed from Fig. 3, ARPA is divided into six separate parts. Nevertheless, Figs. 4 and 5 are divided into four and five separate parts, respectively. It is demonstrated that the method of this paper is more accurate.

In order to better verify the proposed method, we evaluate node importances in Zachary karate club network and 40 HIV/AIDS patients relational network. Fig. 6 indicates the calculation results by using the method of this paper. We select the Top5% ranked important nodes in Zachary karate club. The larger the nodes , the more important. We can see that the

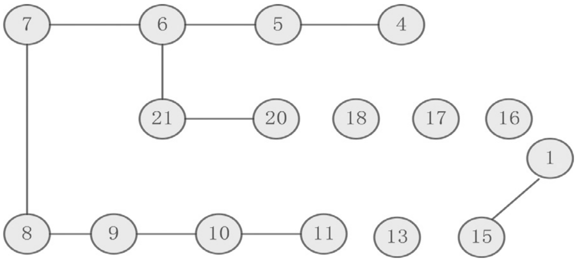

Fig. 3. A part of ARPA after removing the Top5% nodes and connected edges of ARPA by the proposed method.

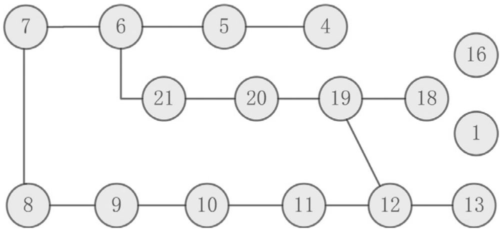

Fig. 4. A part of ARPA after removing the Top5% nodes and connected edges of ARPA by the method in the literature (Zhou et al., 2012).

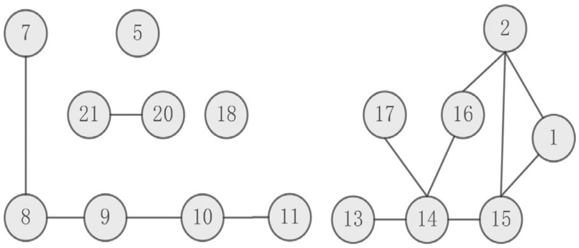

Fig. 5. A part of ARPA after removing the Top5% nodes and connected edges of ARPA by the method in the literature (Zhao et al., 2009).

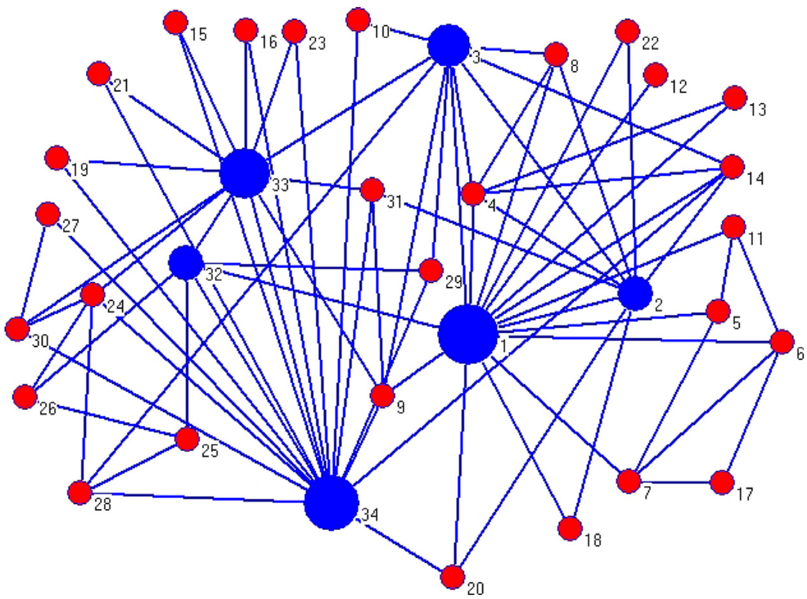

Fig. 6. The Top5% nodes in Zachary karate club network. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
The rankings of Top5% nodes in ARPA.

Rank	The proposed method	Literature (Zhou et al., 2012)	Literature (Zhao et al., 2009)
1	3	2	3
2	14	3	12
3	2	15	19
4	12	14	6
5	19	17	4

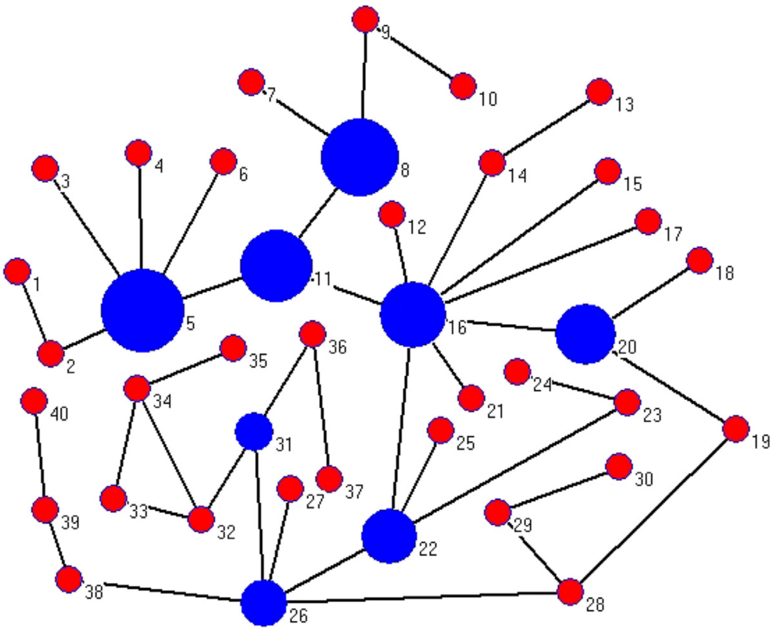

Fig. 7. The Top5% nodes in 40 HIV/AIDS patients relational network. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Top5% nodes can overwhelmingly cover the important nodes in Zachary karate club network. These important nodes are some central figures and they are related to the most popular individuals. They play important roles in one aspect of the whole team.

The proposed method is also applied to a network of 40 homosexuals with AIDS. We selected the Top5% ranked important nodes in the HIV/AIDS network Fig. 7. The larger the nodes (blue nodes) are, the more important they are. Their ranking is compatible with the ranking which is generated by the real spreading process substantially. The Top5%, like nodes 5, 8, 11 and 16, have high current values and are more likely to be bridging nodes. The blue spreading nodes are the most "popular" and would cover almost all the network. At the same time, the blue nodes are considered to be more central than other nodes and they can spread the information (or disease) faster. Moreover, being located in the middle of the contact network, individual 22 is now considered to be the most central. It is rather intuitive to establish that individual 26 is more important while individual 5 is rather peripheral. Therefore, the proposed method can rank the nodes spreading ability correctly.

5. Conclusion and discussion

In this paper, we first reviewed five indicators to evaluate node importance in actual complex networks and analyzed their characteristics. The paper selects these indicators, respectively, from the points of local attributes, global attributes of network, information dissemination and network location. A multi-attribute ranking method is proposed to evaluate node importance comprehensively in this paper. On the whole, our proposed method is verified to be feasible and outperforms other methods by comparative experiments in two actual complex networks. According to the proposed method, we can provide decision support for emergency management in actual engineering, which is overwhelmingly important to national stability and development. In addition, the generalization of the results to any undirected relational network is straightforward.

Please cite this article as: Y. Yang, G. Xie, Efficient identification of node importance in social networks, Information Processing and Management (2016), http://dx.doi.org/10.1016/j.ipm.2016.04.001

Additional Information and Declarations

Funding

This work was supported by the Fund Project of Taiyuan City Science and Technology Special Talents (Grant No.120247-28), and the National Natural Science Foundation of China (Grant No. 61402319). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Yunyun Yang conceived and designed the research, performed the experiments, analyzed the data, wrote the paper, prepared figures and tables, performed the computation work.
- Gang Xie conceived and designed the research, analyzed the data, reviewed drafts of the paper.

Acknowledgments

We thank Z.M. Gao for his valuable comments and useful feedback on the manuscript. Additionally, we particularly thank Y.P. Liang with expertise in technical English editing for the English of this manuscript being improved.

References

Ahlswede, R., Cai, N., Li, S. Y. R., et al. (2000). Network information flow. *Information Theory, IEEE Transactions on, 46*(4), 1204–1216.
Arulselvan, A., Commander, C. W., Elefteriadou, L., et al. (2009). Detecting critical nodes in sparse graphs. *Computers & Operations Research, 36*(7), 2193–2200.
Bhola, V., Grover, M. K., Sinha, M., et al. (2010). Identifying key players in a social network: Measuring the extent of an individual's neighbourhood connectivity. In *Business applications of social network analysis (BASNA), 2010 IEEE international workshop on. IEEE* (pp. 1–6).
Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 1170–1182.
Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology, 2*(1), 113–120.
Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory, 12*(1), 21–34.
Burt, R. S. (2009). *Structural holes: The social structure of competition*. Cambridge, Massachusetts, and London, England: Harvard university press.
de Camargo, J. B. Jr., Neto, M. S., & Pires, S. R. et al. (2012). The role of logistics services providers in The supply chain management: The social network perspective.
Chen, H., & Wang, Y. T. (2012). Threshold-based, heuristic.algorithm., for influence maximization. *Journal of Computer Research and Development, 49*(10), 2181–2188.
Estrada, E., & Rodriguez-Velazquez, J. A. (2005). Subgraph centrality in complex networks. *Physical Review E, 71*(5).N1539-3755.
Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215–239.
Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks, 13*, 141–154.
Guo, S. Z., & Lu, Z. M. (2012). *Complex networks foundation theory* (pp. 250–258). Beijing: Scientific Press.
Hawick, K. A., & James, H. A. (2007). Node importance ranking and scaling properties of some complex road networks. *Research Letter in the Information and Mathematical Sciences, 11*, 23–32.
He, N., Li, D. Y., Gan, W. Y., & Zhu, X. (2008). Mining vital nodes in complex networks. *Computer Science, 34*(12), 1–5.
Hewett, R. (2011). Toward identification of key breakers in social cyber-physical networks. In *Systems, man, and cybernetics (SMC), 2011 IEEE international conference on. IEEE* (pp. 2731–2736).
Hu, J., Wang, B., & Lee, D. (2010). Evaluating node importance with multi-criteria. *Proceedings of the 2010 IEEE/ACM Int'l conference on green computing and communications & Int'l Conference on Cyber, Physical and Social Computing* (pp. 792–797). IEEE Computer Society.
Kitsak, M., Gallos, L. K., & Havlin, S. (2010). Identification of influential spreaders in complex networks. *Nature Physics, 6*(11), 888–893.
Kossinets, G., & Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science, 311*(5757), 88–90.
Liu, J. G., Ren, Z. M., Guo, Q., & Wang, B. H. (2013). Node importance ranking of complex networks. *Acta Physica Sinica, 62*(17), 178901.
Liu, Y., Tang, M., Zhou, T., et al. (2015). Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition. *Scientific reports*, 5.
London, A., Németh, T., Pluhár, A., et al. (2015). A local pagerank algorithm for evaluating the importance of scientific articles. *Annales Mathematicae et Informaticae, 44*, 131–140.
Poulin, R., Boily, M. C., & Masse, B. R. (2000). Dynamical systems to define centrality in social networks. *Social Networks, 22*(3), 187–220.
Ren, X. L., & Lv, L. Y. (2014). Review of ranking nodes in complex networks. *Chinese Science Bulletin, 13*, 4–7.
Ren, Z. M., Liu, J. G., Shao, F., et al. (2013). Analysis of the spreading influence of the nodes with minimum K-shell value in complex networks. *Acta Physica Sinica, 62*(10), 108902.
Stephenson, K., & Zelen, M. (1989). Rethinking centrality: Methods and examples. *Social Networks, 11*(1), 1–37.
Sun, R., & Luo, W. B. (2012). Review on evaluation of node importance in public opinion. *Application Research of Computers, 29*(10), 3606–3608.
Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature, 512*(7513), 126–129.
Wang, L., & Zhang, J. J. (2006). Centralization of complex networks. *Complex System and Complex Science, 3*(1), 13–20.
Wang, X. F., Li, X., & Chen, G. R. (2012). *Network science: An introduction* p. 161. Beijing: Higher Education Press.
Wasserman, S. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.
Xie, X., Li, Y., Zhang, Z., et al. (2015). A joint link prediction method for social network. *Intelligent computation in big data era* (pp. 56–64). Berlin Heidelberg: Springer.
Yu, H., Liu, Z., & Li, Y. J. (2013). Key nodes in complex networks identified by multi-attribute decision-making method. *Acta Physica Sinica, 62*(2), 020204.

Zhang, J., Xu, X. K., Li, P., et al. (2011). Node importance for dynamical process on networks: A multiscale characterization. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 21*(1), 016107.
Zhao, Y. H., Wang, Z. L., Zheng, J., & Guo, X. J. (2009). Finding most vital node by node importance contribution matrix in communication netwoks. *Journal of Beijing University of Aeronautics and A, 35*(9), 1076–1079.
Zhou, X., Zhang, F. M., Li, K. W., et al. (2012). Finding vital node by node importance evaluation matrix in complex networks. *Acta Physica Sinica, 61*(5), 050201.
Zhu, Y., Meng, Z. Y., & Kan, S. Y. (1999). Determination of weight value by AHP. *Journal of Beijing Jiaotong University, 23*(5), 119–122.