Soft Comput (2018) 22:453–464
https://doi.org/10.1007/s00500-016-2335-3

METHODOLOGIES AND APPLICATION

Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm

Li Yang[1] · **Yafeng Qiao**[1] · **Zhihong Liu**[1] · **Jianfeng Ma**[2] · **Xinghua Li**[2]

Published online: 12 September 2016

Abstract In online social networks, there are some influential opinion leader nodes who can be used to accelerate the spread of positive information and suppress the diffusion of rumors. If these opinion leaders can be identified timely and correctly, there will be contributing to guide the popular opinions. The closeness is introduced for mapping the relationship between the nodes according to the different interaction types in online social network. In order to measure the impact of the information transmission between non-adjacent nodes in online social networks, a closeness evaluating algorithm of the adjacent nodes and the non-adjacent nodes is given based on the relational features between users. By using the algorithm, the closeness between the adjacent nodes and the non-adjacent nodes can obtained depending on the interaction time of nodes and the delay of their hops. Furthermore, a more accurate and efficient betweenness centrality scheme based on the optimized algorithm with the degree of closeness and the corresponding updating strategy. The opinion leader nodes should be identified more accurately and efficiently under the improved algorithm because the considering of closeness between nodes in the network. Finally, the maximum spreading experiment is done for comparing the proposed method with other existing identifying opinion leader selecting schemes based on the Independent Cascade Model. The result of experiment shows the effectiveness and practicality of the evaluating algorithm.

Keywords Social networks · Closeness · Independent cascade · Opinion leader nodes

Communicated by V. Loia.

✉ Li Yang
yangli@xidian.edu.cn

Yafeng Qiao
yfqiao@mail.xidian.edu.cn

Zhihong Liu
zhliu@mail.xidian.edu.cn

Jianfeng Ma
jfma@mail.xidian.edu.cn

Xinghua Li
xhli@mail.xidian.edu.cn

1 School of Cyber Engineering, Xidian University, Xi'an 710071, China

2 Shaanxi Key Laboratory of Network and System Security, Xi'an 710071, China

1 Introduction

Identifying the most influential nodes in a complex network is an important step toward optimizing the available resources and ensuring the more efficient of information spreading (Kitsak et al. 2010; Qin et al. 2016). Online social network is very popular today with the rapid development of Internet and cloud computing (Li et al. 2015). In social networks, messages are spread between the nodes. The most influential nodes have become an important factor which will affect the diffusion degree of the messages (Putzke and Takeda 2016). We all know that in social networks the most influential nodes have some characteristics. Firstly, they have abundant information resources. Secondly, they prefer to express their views and their discourse are thoughtful or critical. Thirdly, they have a strong ability on the media. Therefore, in social networks these influential nodes can also be named as opinion leader nodes. We can use the opinion leader nodes to expand the spreading range of positive news, so that more users in the social networks can receive the true message (Zhang and Wang 2013; Wang and Ma 2015). Moreover, the opinion leader nodes can be used to restrain the spread of public opinion by influencing on them. We can take advantage of these two characteristics in a social network to achieve the appropriate purpose, such as spread the correct message more efficiently or guide the spreading of the public opinions to a certain extent. Therefore, in order to control the spread of messages in social network more better, either for the purpose on accelerating the proliferation of positive news or suppressing the public opinions, there is need to identify the opinion leader nodes in social networks accurately.

In most social networks, users' choice will usually be affected by their friends, and such influence is greater than other non-friend nodes. The model proposed in Granovetter (1973) shows that the different linking strength played a different role in the process of information spreading. The strong links will usually form high level of trust relationship with other nodes and disseminate much information, while the weak links are just in the opposite. Bakshy (2009) and Onnela (2007) proved that the above-mentioned results with experiments, and they concluded that the information spread by local links, named as acquaintances spread. Therefore, a social network shows the influence of "word of mouth effect". By "word of mouth effect", we can seen that the more closely between users, the more easily acceptance and dissemination the recommended message will be. Obviously, the speed of news spreading is proportional to the relationship between nodes. In social networks, if a node has a higher closeness with other nodes, it will have a higher influence. Therefore, we can identify the opinion leader nodes by measuring the closeness between the social network nodes.

The paper is mainly aimed at the identification of the opinion leaders in online social networks. The main contributions are as follows. Firstly, different interaction types are analyzed in online social network, then we define them as different closeness for mapping the relationship between the nodes. Secondly, a new algorithm of the closeness between neighbor nodes and non-adjacent nodes in social networks is introduced. The closeness between nodes is evaluated by the proposed algorithm. The social network model will be bordered and updated depending on the nodes closeness, and the nodes closeness is set as the weight of the link. Then according to the nodes closeness and based on the algorithm of betweenness, the opinion leader nodes in social networks can been identified. Thirdly, the experiments according our algorithm are given finally. In the experiments, we set the opinion leader nodes as the initial seed nodes and evaluate the maximum value of infected nodes by using the Independent Cascade Model (short as IC Model) to spread the message. By the diffusion effect, the effectiveness of opinion leader nodes should be determined.

The rest of the paper is organized as follows: Sect. 2 is the related works. Section. 3 shows the relevant definitions of the proposed scheme. Section. 4 introduces the definition of closeness and the influence factors. Section. 5 gives the new algorithm of the opinion leaders identifying. In Sect. 6, the experiment and the analysis is given. In Sect. 7, the conclusion and some ideas on the future works will be presented.

2 Related works

The measurement of the centricity index and the information propagation model are two important concepts in our scheme. The centricity index is used to identify opinion leader nodes on the basis of the closeness between nodes. The maximize information diffusion experiments using will be done in the information model by using opinion the leader nodes as a set of seeds. Then we can evaluate the effectiveness of opinion leader nodes accurately based on the maximizing diffusion.

2.1 Centrality

There are some benefits to identify the influential nodes in complex networks, such as understanding the dynamics and optimization the design of the influence network or finding the vulnerability of network (Holme et al. 2002). It is also helpful for controlling the information flow in the networks. For identify the key nodes in networks, some central characteristic index has been proposed; they are degree centrality, closeness centrality (Sabidussi 1966), betweenness centrality (Freeman 1978) and eigenvector centrality (Bonacich 2007) and so on.

The advantage of the degree centrality is simplicity, and the index value of a node can be calculated only based on the local topology. But in degree centrality method, the global topological properties are ignored. Even if a node is connected with many nodes, but it does not be in a position where can be fast accessing information, then it can not be a key node. The closeness centrality is defined as the reciprocal sums of the shortest path from one node to all other nodes in the network. But the closeness centrality can not be applied to the networks with two disconnected components. Although the global topology is considered in betweenness centrality algorithms, it also can be applied to the networks with two disconnected components, the nodes who is not on any shortest path are ignored. If two or more different components exist, the eigenvectors of adjacency matrix can only recognize one component. It means that the eigenvector centrality is not suitable for analyzing the networks which contains multiple components.

By compared the betweenness centrality algorithm with the degree centrality algorithm and the closeness centrality algorithm, the results show that betweenness centrality algorithm can identify the key nodes more accurately. And the betweenness algorithm offers more advantages than eigenvector centrality when the network has an asymmetric adjacency matrix. Because the relationship between the social network users is directed, the betweenness centrality algorithm is used as a basic algorithm to construct our algorithm for identifying the opinion leader nodes.

In betweenness centrality algorithm, the nodes' betweenness is counted by the number of shortest paths passing the node. The betweenness centrality is mostly applied to the unweighted undirected graph initially. Then a more general framework for calculating the betweenness of the weighted directed graph is set up in Brandes (2008). In the further research, some more effective and stable algorithm is given (Segarra and Ribeiro 2014). The algorithm has been widely used in complex networks research works such as analysis of social networks (Said et al. 2008), identifying critical nodes in a wireless network (Guo et al. 2014; Shen et al. 2015) and studied the activity and the importance of mobile telephone network nodes (Catanese et al. 2013).

The computational overhead of the betweenness centrality algorithm is costly. The algorithm proposed in Brandes (2001) can be used to calculate the betweenness of all vertexes in the graph accurately. A random algorithm is introduced in Bader et al. (2007) for the betweenness centrality evaluation. The calculation and the time costs in this method increase greatly with the size of the network, while the accuracy reduces than the previous. The opinion leader nodes identifying method is designed based on the algorithm in Domingos and Richardson (2001) in our scheme.

2.2 Propagation models

Several information propagation models can be used to describe the propagation of information in social networks, including the Epidemic model, the Linear Threshold Model (short as LT model) and the Independent Cascade Model (short as IC model). A general assumption in these models is that uninfected nodes can be infected by infected nodes in a certain probability. The process is one type of the cascading affect. The number of nodes to be infected will be in a stable state when the information is spread to a certain extent. This state represents the maximize degree of the information diffusion.

Domingos and Richardson studied on the issue of how to maximize information diffusion in complex networks firstly in Domingos and Richardson (2001) and Richardson and Domingos (2002). They gave the definition and the evaluation indicators of the influence. They also gave a probability propagation model for nodes selection when messages propagation in social networks. On the basis of this definition, Kempe and Kleinberg (2003) summarized the problem of information propagation as finding a set of nodes which have the maximize diffusion area. At the same time, they presented a greedy algorithm to solve this problem. Saito (2008) made a detailed description on the effects of the propagation model under the influence maximization problem, named as Linear Threshold model and Independent Cascade model. The two models have been the basic models on studying the influence maximization in social networks.

Since each infection process in the IC model is independent, the activation relationship between the node and its outgoing edge neighbor is only considered, while the effects of its incoming edge neighbor are completely disregarded. IC model is more aligned with the social networking application background than LT model. In our scheme, the validity of opinion leader nodes is tested by IC model.

3 Related definitions

3.1 Betweenness centrality

Betweenness is a centrality measurement based on shortest paths, including the node betweenness and the edge betweenness, which is widely used in complex network analysis. The edge betweenness is the proportion of the number of the shortest paths through the edge to the total number of all shortest paths. The node betweenness is the proportion of the number of the shortest paths through the node and the total number of all shortest paths. The betweenness centrality we used here means the node betweenness. Its definition is shown in Eq. (1) (Freeman 1977).

$$C_B(x) = \frac{2\sum_{j<k} g_{jk}(x)}{(n-1)(n-2)g_{jk}} \tag{1}$$

where g_{jk} represents the number of the shortest paths between node j and node k. $g_{jk}(x)$ represents the number of the shortest paths through the node x between node j and k. Then the maximum possible node betweenness will be $(n-1)(n-2)/2$.

3.2 Independent cascade model

Independent cascade model (IC model) is widely used in the social networks for studying the influence of the certain node (Kempe et al. 2003). In IC model, each node has three states, namely non-infected, infected and contagious. When the node is in the contagious state, it is also in the infected status and can disseminate information. Given a weight value for each edge $(u, v) \in E$ in graph $G = (V, E)$, the weight presents the transmission probability $pp(u, v)$ between nodes. And the probability $pp(u, v)$ means that when a node u in the uninfected state will be infected by a node v at the next time with the probability $pp(u, v)$.

The Independent cascade process of IC model is as blow:

Given a set of nodes $S \subseteq V$, all the nodes in the set are in the contagious state. For other nodes $u \notin S$, they are all in the non-infected state. According to the following process, we can maximize the diffusion:

(1) $S_t \subseteq V$ is the set of nodes which are infected in step t, where $t \geqslant 0$ and $S_0 = S$;
(2) In step $t + 1$, each node can infect its neighbor node $v \in V \backslash U_{0 \leqslant i \leqslant t} S_i$ by the probability $pp(u, v)$;
(3) The process will stop in the step t while $S_t = \varnothing$.

In IC model, a node only has one chance to infect its neighbor nodes according to the above process. If the node i has infected in step t, it will infect its neighbors in step $t + 1$ by a probability $pp(u, v)$.

4 Closeness between nodes

In a social network, it has a corresponding degree of closeness between any two users. The degree of closeness between two users describes their relationship of two users. There is need to find a way to make the closeness more meaningful. We design the following procedure to assess the degree of closeness between users. We analyze the influence factors which affect the degree of closeness between the nodes and use the appropriate variables present these influence factors. Then the value of closeness will be comprehensive evaluation of these factors according to the improved algorithm.

4.1 Influence factors of closeness

In social networks, the relationship between nodes can mutual mapping with users in reality. Users interact with each other in various ways. According to the users' classification of friends, list can show their relationship clearly. A variety of social networks default buddy type and interaction type statistics as shown in Tables 1 and 2, respectively.

Based on Tables 1 and 2, the buddy types and interactive ways of the social network node can be summarized as below:

(1) Buddy types are families, friends, classmates, colleagues, public homepage and public figures;
(2) Interactive ways are private message, release status, comments and chatting.

In social networks, the buddy type of users can reflect the closeness between users and different relation users will use the different interaction type. The close friends prefer to send private messages or chatting usually, while normal friends will give each other a comment more often. Then we can measure the closeness between nodes effectively depending on the counting of the interaction of different types with different friends.

The interaction between users in social networks is with the following characteristics. If you maintain a close relationship with others, you have to pay the cost of the corresponding time. It manifest as the frequency of forwarding messages or comments. In this paper, we assume that there is no such phenomenon that the closeness between two nodes is higher but with low frequent interaction.

Since we can not get the interactive situation of private message, chatting and so on, the mention network is used. In order to reduce the complexity of the data collection, the following factors are used to measure the closeness between nodes.

(1) The average time gap of the interaction in a moment is used instead of the count of interaction. The smaller the interaction gap is, the more often of the interactive and there will be more often closeness of two users will be.
(2) The interaction between nodes is simplified. The interactive mode of nodes can be get without knowing the user's privacy, like a mention network.
(3) The buddy type of nodes is simplified to friends or strangers in this situation, where friends mean the nodes which are families, friends, classmates and colleagues. Strangers include the public homepages and the public persons followed by users and the nodes do not focus on each other. In a word, if nodes focus on each other, they are friend, otherwise they are stranger. For example, in Sina Weibo and Twitter only friend nodes follow each other, public accounts seldom follow each other.

In our model, the buddy type and the interaction way are simplified, and the accuracy of closeness measuring between nodes will be reduced; the difficulty of data collection and computational complexity will be reduced too. However, the time gap and the frequency of interaction can cover the shortage of the simplified above.

4.2 Closeness between nodes

The users and their relationship in social networks can be mapped to a graph model with nodes and links in the process of quantification. We first create a social network model with the set of nodes V and the set of directed links E. This model is a directed graph $G = (V, E, W)$. The set of users in social networks mapped to the set of nodes V, and the relationship between users mapped to the set of directed links E. The value of closeness between users mapped to the weight W.

According to the effect factors described above, we will calculate the degree of closeness between the nodes with the following three aspects.

Table 1 Common buddy types in social networks

Socials	Buddy types					
	Ordinary	Intimate	Families	Classmates	Homepages	Public persons
QQ	√	√	√	√	–	–
Facebook	√	√	√	–	√	√
Sina Weibo	User self-classification					
Twitter	User self-classification					

Table 2 Interaction types in social networks

Socials	Interaction types			
	Private messages	Something new	Comments	Chat
QQ	–	√	√	√
Facebook	√	√	√	–
Sina Weibo	√	√	√	√
Twitter	√	√	√	–

4.2.1 Time interval of interactions

Assuming that the interval time of interactions between two nodes is t. By calculating the average of interactions within the time period T, we can avoid the phenomena of closeness inaccurate measurement which is caused by unexpected interaction between two nodes.

The average interval time is defined as Δt, and it can be calculated with Eq. (2).

$$\Delta T = \frac{1}{n-1}\sum_{i=1}^{n-1} \Delta t_i \tag{2}$$

where n is the interaction times of the two nodes within time T, Δt_i is the interval of interactions between ith time to $(i+1)$th time.

4.2.2 Interaction types

Some public accounts in social networks will continually push messages to users or release new states frequently. But we can not say that the users and these public accounts are close in this situation. We set different weights for different types of nodes to dealing with this situation. Then the assignment of interaction time is defined by Eqs. (3) and (4).

$$w_s = \alpha \Delta T \tag{3}$$

$$w_f = \beta \Delta T \tag{4}$$

where w_s is the weight of the link when the relations of two nodes are strangers, w_f represents the weight of link when the relations between two nodes are friends. α and β are two balance factors which will be discussed later.

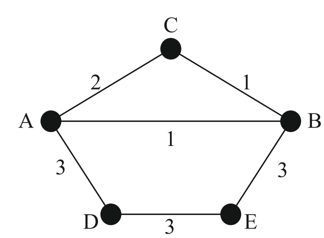

Fig. 1 Effects of the node closeness and time delay in hop

4.2.3 Hops between nodes

In social networks, there is no direct link of two nodes does not mean that there is no closeness. As shown in Fig. 1, the weight of a link represents the closeness of any two nodes. So we can assume that the larger the weight is then the greater the closeness will be. Because the nodes D and E, the nodes E and B both have a greater closeness; the news released from node D to the Node B through the node E will be more quickly, although there is no direct link between node D and B. It shows two nodes which are not adjacent but have closeness.

Since there is a delay in each hop when massages spreading between nodes. As shown in Fig. 1, the message can be translated from node A to B. The closeness between node A and B is very small. Although the weight of link $A- > D- > E- > B$ is larger, it is a multi-hop path. So we need to check the closeness and the delay of hops in path $A- > D- > E- > B$ and path $A- > B$. But the transmission speed of two paths can not be compared directly.

In the process of information dissemination, the existed work (Lu et al. 2014) shows while the information starting translate from one node, then after three to four hops its influence will be very weak. In our scheme, we only consider the closeness between nodes whose interval is within three hops.

In online social networks, the process of message spreading between nodes is independent to each other. The speed of propagation and the degree of acceptance of one node can not be effected by its previous hop nodes. As shown in Fig. 1, the message m is created in node A and then be send from A to B via C; the communication process from node A to C and the process between C and B are independent to each other.

Thus, we can calculate the closeness between the nodes as shown in Eq. (5).

$$w_{i,j} = \gamma_1 \sum_{n=1}^{h_{i,j}} w_n + \gamma_2 h_{i,j} \tag{5}$$

where γ_1 and γ_2 are the weight of the interaction time and the hops, respectively, and $\gamma_1, \gamma_2 \leqslant 1$. $h_{i,j}$ represents the hops from node i to node j and $h_{i,j} \leqslant 3$. w_n represents the weight which the node be away from node i with n hops and the node be away from node i with $n+1$ hops.

After comprehensive quantization, we get the closeness of any two nodes in the online social network. The smaller of the weight value means the more frequently of the interaction between nodes, while their closeness is in a higher degree correspondingly.

5 Identify the opinion leader nodes

We can identify the opinion leader nodes based on the betweenness centrality by calculating the degree of closeness between the nodes and making it as the link weights in the directed graph. Although betweenness centrality algorithm is widely used, there are still some differences in our scenarios.

(1) The main idea of the betweenness is based on the statistics of shortest paths through the node v. If it is a weighted graph, the shortest paths are the paths which has the minimum sum of weights. In social networks, the process of message spreading is independent to each other, so the weight can not be added directly.
(2) Only the weight is considered when calculating the betweenness in the weighted graph, the hops of nodes do not considered. In the application of the network environment, any nodes which receiving and transmitting the news will pay the time. The hops between nodes will cause the time delay.

In order to make up for the shortage of betweenness centrality when use it based on the closeness, we take the following processes to improve the betweenness centrality algorithm.

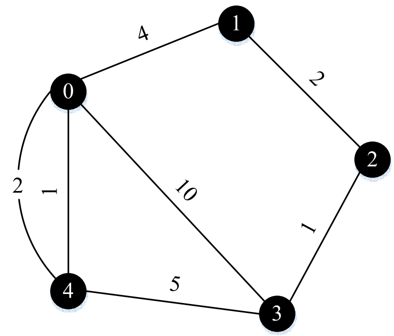

Fig. 2 Calculate the closeness between nodes

(1) Given the direct weighted graph, calculate the closeness between node i and all other node which distance from node i within three hops by Eq. (5).
(2) If there is not only the direct link P_d between node i and j, but also exists another path P_h. The closeness w_d and w_h obtained from these paths should be compared, respectively. If $w_d > w_h$, the weight of P_d with w_h is updated. Otherwise the original graph is unchanging.
(3) If there is no direct link P_d between node i and j, only exists another path P_h. Add a link from node i to j and set the weight of the link as $\mathrm{Min}(w_h)$.
(4) Update the graph and calculate the betweenness for the new graph.

As shown in Fig. 2, the weight between nodes represents the interaction interval of nodes. The smaller the weight is, the more frequently of the interaction between nodes will be.

In Fig. 2, we assume that only the bi-directional links between nodes are friend relationship. Otherwise, they are strangers. The algorithm of identifying opinion leader nodes can be shown as Algorithm 1.

We can calculate the betweenness and identify the opinion leader nodes by the output of Algorithm 1. We illustrate the main process of Algorithm 1 by Fig. 2.

First, the closeness between the node and its direct neighbors, that is one hop, is calculated in Fig. 2. Then the closeness as two and three hops from node 0 is computed. The results between node 0 and node 2 is $W_{(0,2)} = (4+2)\alpha\gamma_1 + 2\gamma_2$, between node 0 and node 3 is $W_{(0,3)} = (4+2+1)\alpha\gamma_1 + 3\gamma_2$.

We give the closeness between node 0 and node 2, 3, when α,γ_1 and γ_2 are taken part of the value in Table 3.

When $\gamma_1 = 0.7$, $\gamma_2 = 0.3$, $\alpha = 2$, the closeness between Node 0 and Node 2 is 5.0, and the closeness with Node 3 is 6.3. Because there is a direct link between Node 0 and Node 3, while the weight $6.3 < 10$, then the weight of path $0 \rightarrow 3$ should be updated as 6.3. Add a path from $0 \rightarrow 2$ and

Algorithm 1 Updating algorithm based on the closeness

Input:
An adjacency matrix whose weight is closeness;
Nodes : All the nodes in the graph;
Neighbors : Nodes which distance from node i within 3 hops;
$h_{i,j}$: The hops between node i and j;
$dw_{i,j}$: The weight of a direct link between node i to node j;
Output:
An adjacency matrix after update.
for $i \in$ Nodes **do**
  for $j \in$ Neighbors **do**
    $w_{i,j} = \gamma_1 \sum_{n=1}^{h_{i,j}} w_n + \gamma_2 h_{i,j}$
    if there is a direct link between i and j **then**
      if $w_{i,j} < dw_{i,j}$ **then**
        $dw_{i,j} \leftarrow w_{i,j}$
      else if Add a link from i to j **then**
        $dw_{i,j} \leftarrow w_{i,j}$
      end if
    end if
  end for
end for
return

Table 3 Influence of different balance coefficient on intimacy

Values		γ_1 0.1, γ_2 0.9	γ_1 0.3, γ_2 0.7	γ_1 0.5, γ_2 0.5	γ_1 0.7, γ_2 0.3	γ_1 0.9, γ_2 0.1
$\alpha = 2$	$W_{(0,2)}$	3.0	5.0	7.0	9.0	11.0
	$W_{(0,3)}$	3.1	6.3	8.5	10.7	12.9
$\alpha = 3$	$W_{(0,2)}$	3.6	6.8	10	13.2	16.4
	$W_{(0,3)}$	4.8	8.4	12	15.6	19.2

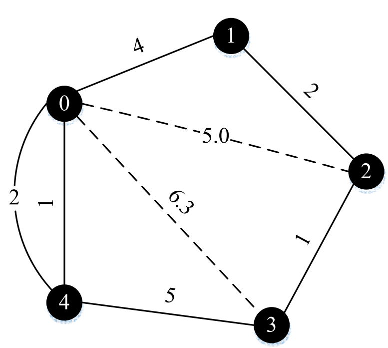

Fig. 3 The updated graph

set its link weight as 5.0. According to the above calculation results, the graph in Fig. 2 is updated too, and the results is as shown in Fig. 3.

In the new algorithm, a process of edge adding is introduced in network on the basis of the betweenness centrality. The time complexity of this process is increasing while the nodes increasing, and it will cause very little impact on the overall time complexity to the recognition algorithm.

6 Experimental analysis

6.1 Modeling

The real-world networks have a shorter average path and the larger clustering coefficient overall. In fact, the real-world networks have the characteristic of small world. In online social networks, every user has some friends. There is only one hop between the user and his/her friends. But there are minority nodes which connect to a large number of other nodes in local, and many nodes have fewer links. It means that there are little hub nodes in a small group. These hub nodes and other nodes constitute the scale-free properties. Then we build a small-world network model with scale-free properties to match the situation.

The WS small-world model, proposed by Watts and Strogatz (1998), used a strategy of random reconnection the graph of the rules. This method can destroy the connectivity of the graph. The NW small-world network is proposed by Newman and Watts (1999) after WS model. The NW model build the network by adding edge randomization. Our network model is based on the NW model.

The scale-free network has two important features: 1) Increase features: The scale of the network is growing. 2) Preferential connection characteristic: The new nodes tend to connect to the nodes with higher degree. It is the phenomenon of the rich richer and the poor poorer. Social networks also have both of the two features of the scale-free networks. For example, there are plenty of new user registration account and use the social networks. In Weibo and Twitter, the new users will usually give precedence to the public counts or celebrities and the friends he knew well.

According to the NW small-world model construction algorithm and two properties of scale-free networks, the small-world network model with scale-free properties construction algorithm may be described as follows:

(1) Starting with the rules graph, consider a nearest-neighbor coupled network with N nodes. They surrounded as a ring, and every node connects with its each $K/2$ neighbor node around itself.
(2) Randomization adding edges. Select a pair of nodes with the probability p randomly and add a new edge between them. Between any two different nodes there is one edge at most, and each node can not connect with its own.
(3) Increase features. After the process of randomization, one new node is introduced each time, and the new node is connected to m existing nodes, and $m \leqslant N$.
(4) Preferential connection characteristic. A new node connects to an old node i with the probability $\prod_i = \frac{k_i}{\sum k_j}$.
(5) Preferred to connect with friends. The users are more likely to build relationship with his friends in social net-

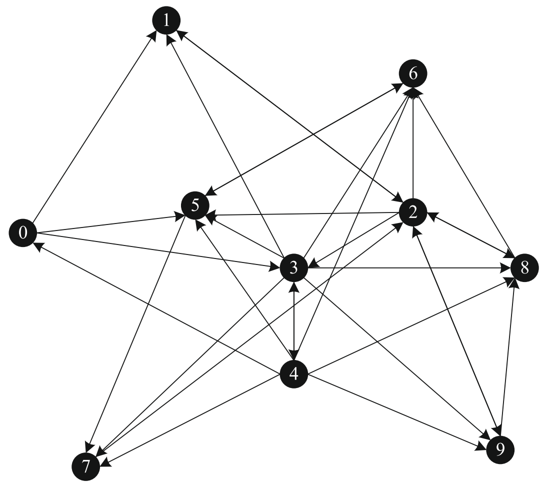

Fig. 4 Small-world network with scale-free characteristics

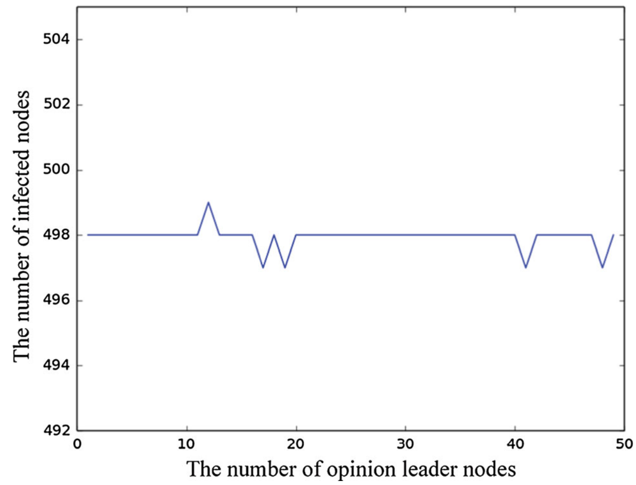

Fig. 5 Different number of seed nodes

works. But a model network, it is difficult to judge the friendship between certain nodes. In the process of model built, $m/2$ nodes are connected as preferential connection, while the other $m/2$ nodes are connected randomly.

The small-world network with scale-free characteristics construction described above is shown in Fig. 4. It is clear that Node 3 has a scale-free characteristics.

6.2 Node influence

6.2.1 Improvement of the IC model

Considering the information dissemination characteristics and the relationship between social network users, IC model has some aspects not applied.

(1) In IC model, the probability infected between two nodes is independent. In social networks, the information spreading of nodes is also independent to each other. But the process of information spreading may cause delay. And the more the hops are, the larger the delay is.
(2) The node in IC model only has once time to infect his neighbor node. In social networks, such as Weibo, when a node tweeting a message, his friends will receive this message when it was flooded in a period of time.

According to the features of social networks, we take the following strategies to improve the above two aspects. When a node i is infected at time t, then node i will keep the infectious during $t + n$. If all the neighbors of node i have been infected at time $t + m$, there is $(m < n)$, node i will be deleted from the infected set of nodes. We then take $n = 3$ in the simulation experiment later. It means that every infected node can not infect the node which distance is greater than 3.

6.2.2 Influence of the opinion leader nodes

The information diffusion experiment is done as taken the opinion leader nodes as the initial seeds of IC model. Then we compare the influence from three aspects, the number of seed nodes, the probability of infection, the effect of step length.

(1) The effect by different numbers of seed nodes

Different number of seed nodes are selected under the same network topology. These nodes are selected from highest to lowest by closeness. The contrast result of diffusion effect is shown in Fig. 5.

According to Fig. 5, a network with 500 nodes can achieve a better diffusion effect when the number of the seed nodes is 1. Because of the probabilistic characteristics of the IC model, the effect has slight fluctuation. And the slight fluctuation maintains within the two nodes.

(2) The effect by different probability of infection

In social networks, the acceptance situation for the message has little difference between nodes. In this part, we compare the effect of different infection probability. The results are shown in Fig. 6.

As Fig. 6 shows, the degree of the maximize diffusion increases with the probability increasing under the condition of invariable seed nodes number. The degree of maximize diffusion will be gradually stabilized when the infection probability reaches a certain value. In Fig. 6, all the 500 nodes are infected when the infection probability reaches 0.6 and

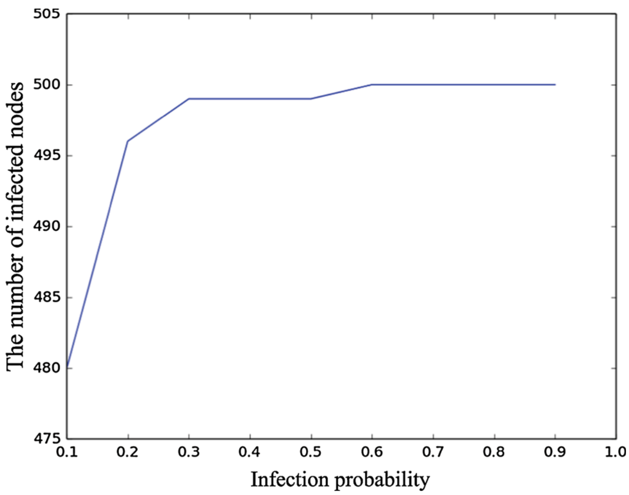

Fig. 6 Different probability of infection

the maximum diffusion get the maximum scale. Then the degree of maximize diffusion will be gradually stabilized even the infection probability is larger than 0.6.

Thus, the information can not be spread when the infection probability is very small. While the infection probability is much bigger, almost every node is infected. If we want to get a better effect in the experiment, an intermediate value should be taken.

(3) The different effect of step lengths

In social networks, the impact of a user to other users will be a continue process. The effect of one message will end until be flooded by many subsequent messages. In the experiment, we compare the diffusion effect of different step lengths. The result is shown in Fig. 7. Where the abscissa indicates the number of nodes in the seed, the ordinate represents the number of nodes be infected.

In Fig. 7a, the step length is 1, which means if a node is infected as time t, it will be infective at time $t + 1$. The step length in Fig. 7b is 2, which means if a node is infected as time t, it will be infective at time $t + 2$. In Fig. 7c is 3, which means if a node is infected as time t, it will be infective at time $t + 3$. As shown in Fig. 7, the size of maximum diffusion has the relationship $(c) > (b) > (a)$. It means that the longer the step is, the bigger the size of information diffusion will be.

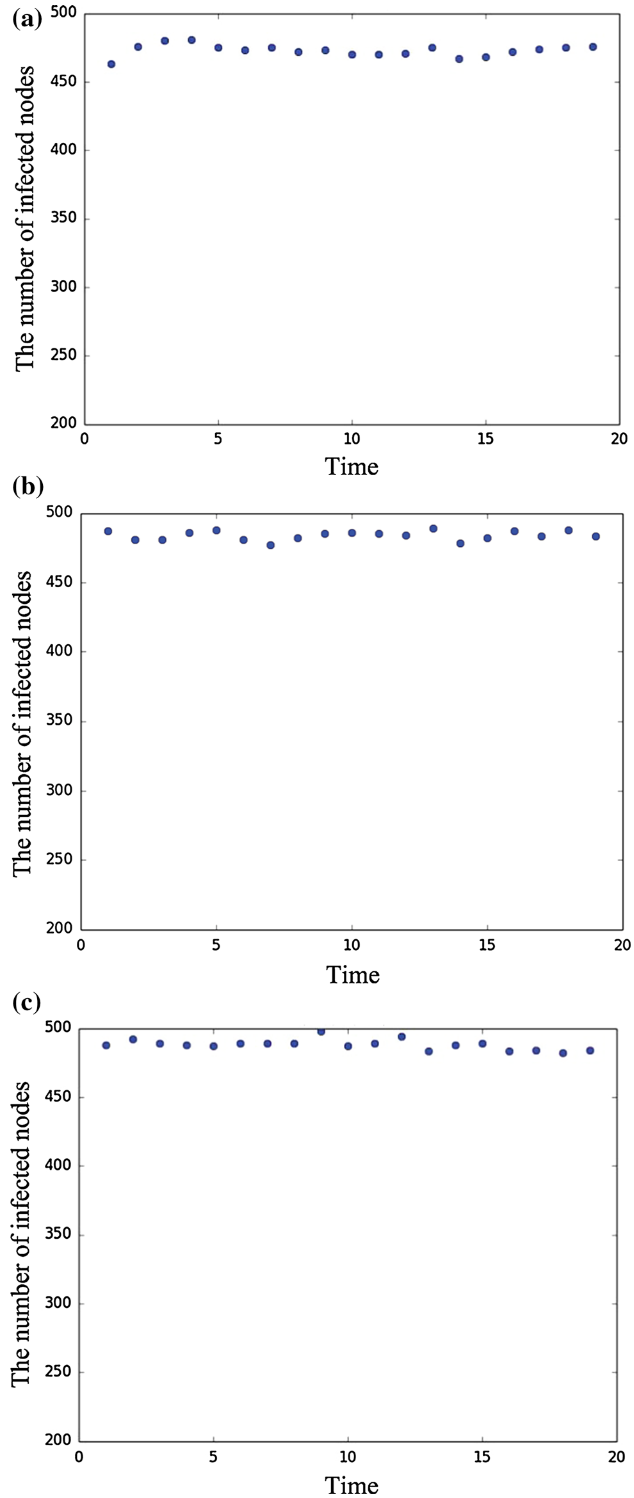

Fig. 7 Different effect of step length. **a** step length = 1, **b** step length = 2, **c** step length = 3

6.2.3 *Availability of opinion leader nodes*

In order to verify the effectiveness of opinion leader nodes, we compare the proposed method with other existing schemes. The compared methods are as blow:

(1) Randomly select N nodes in the network as opinion leaders;
(2) Select N nodes as opinion leaders according to the degree of order from big to small;

(3) Select N nodes as opinion leaders according to the betweenness of order from big to small under betweenness centrality.

In the network with 500 nodes, we identify the opinion leaders in the top ten nodes according to the above method. The results are shown in Table 4, which identify the opinion leaders in the top ten nodes according to four methods.

From Table 4, we can see that besides the method of random selection seed nodes, the other three methods have the phenomenon of repetition in the top ten nodes. So when doing the maximization diffusion experiments under the IC model, part of the nodes which are not repeated just be chosen. And only one node is chosen as the initial node each time. Choosing different nodes in the maximization diffusion will bring more effective experiment results.

The results are shown in Fig. 8. The Node 4, Node 42, and Node 238 are opinion leaders which has been chosen by random selection, degree centrality and closeness method, respectively. The initial process of the closeness method is relatively slow, but it can get the biggest infection nodes when reached the steady state. The reason for slow starting in the

Table 4 Default buddy types and interaction types in social networks

Rank	Random	Degree	Betweenness	Closeness
1	448	238	164	164
2	352	10	75	75
3	290	184	248	248
4	131	43	46	46
5	4	111	184	184
6	261	238	92	92
7	71	108	57	57
8	231	64	10	10
9	329	56	194	42
10	10	136	42	185

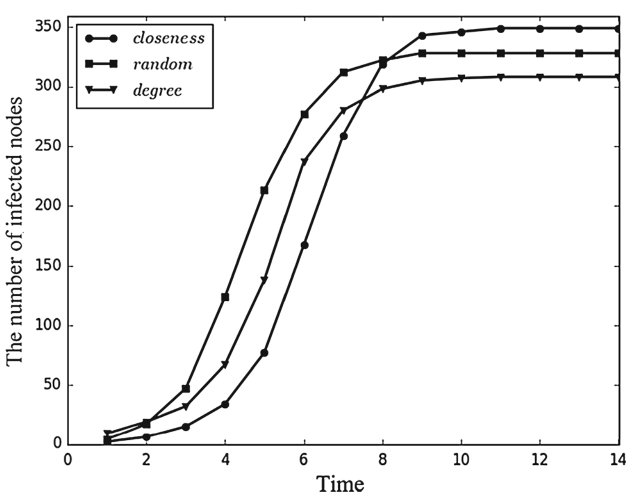

Fig. 8 Compare the closeness with random and degree

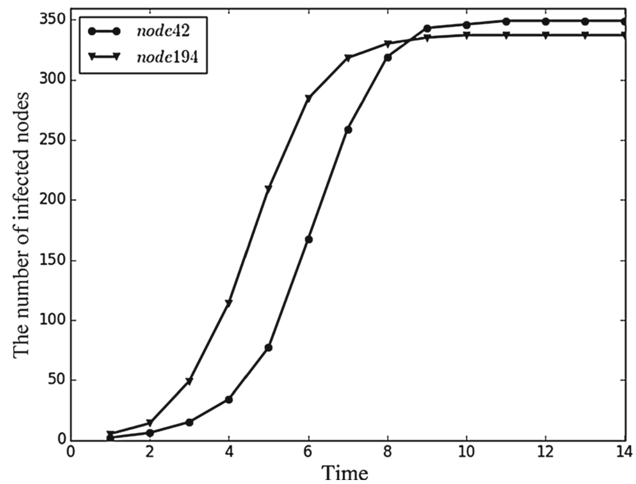

Fig. 9 Compare the closeness with the betweenness

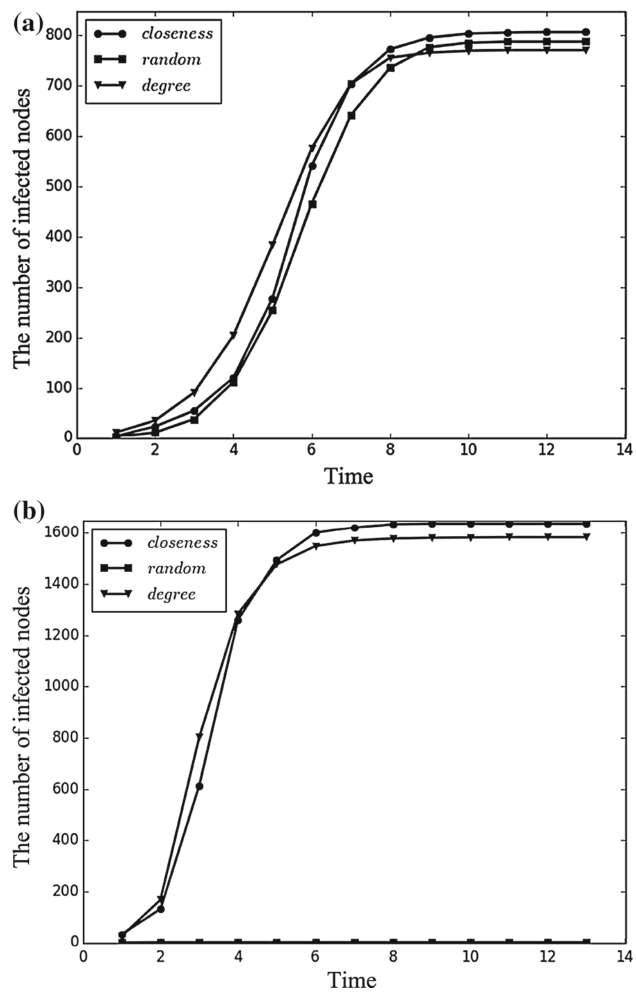

Fig. 10 Compare the closeness with random and degree. **a** 1000 nodes, **b** 2000 nodes

closeness method is the process of the identification of opinion leader nodes. Unlike the degree centrality method, not the large degree nodes but the greater influence neighbor nodes are selected in the closeness method. In the initial spreading process, only a small number of nodes are infected, but it will grow rapidly later.

Because the opinion leader nodes identified under the betweenness method and the closeness method are similar. However, some nodes have a different order; we individually selected two nodes with different orders. The compared results are shown in Fig. 9.

In order to get more fully verifying on the effectiveness, we take the network with 1,000, 2,000, 3,000 nodes to do the maximize diffusion experiments, respectively, and the results are shown in Figs. 10 and 11. The compared results of the proposed method with others are same as the network with 500 nodes. It is clear that the closeness method is more accurate and effective.

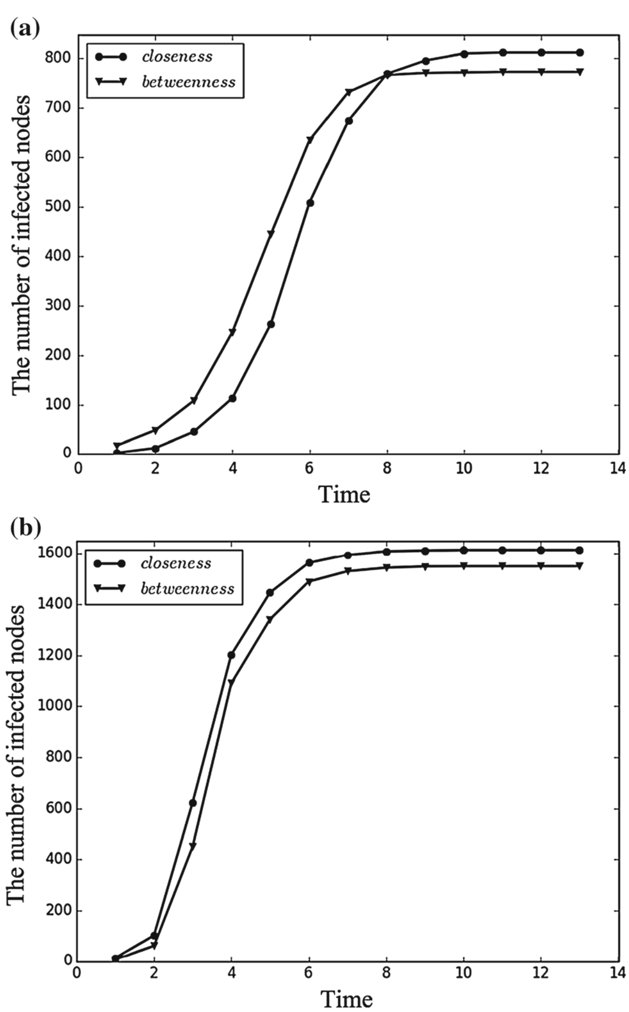

Fig. 11 Compare the closeness with the betweenness. **a** 1000 nodes, **b** 2000 nodes

7 Conclusions

In the paper, a quantitative method of the node closeness is given in social networks. The situation of non-adjacent nodes and the delay of information spread are considered in the method. It is more fit for the social network application environment. The opinion leader nodes are identified by using the betweenness centrality algorithm based on the quantitative method. Finally, the validity of the opinion leader nodes is verified through the IC model.

The experiment results show that the diffusion effect of closeness is better than betweenness centrality algorithm. Although the opinion leaders identified by the closeness algorithm and the betweenness algorithm is similar, we select the different part for experiments. And the opinion leader nodes identified by the closeness algorithm much better than the random and the degree centrality algorithms.

Due to the social networks, each node represents an individual in the real world. And in the process of information transmission, the individual's state is a complex process, so a more fine-grand model should be given to refine the node state. We prefer to take some issue under cloud computing environment (Li et al. 2013, 2014). But we can see that this problem will be still some challenge when put it on the encrypted data (Xia et al. 2015; Fu et al. 2014, 2016). Furthermore, the simulation results are some preliminary when compared with prior solutions, we consider to evaluate the scheme on real-world online social network dataset(s), such as Facebook, Twitter and Youtube as our next step too.

Acknowledgements We would like to thank the anonymous reviewers for their careful reading and useful comments. This work was supported by the National Natural Science Foundation of China (U1405255, 61202390), the China 111 Project (B16037), the Foundation of Science and Technology on Information Assurance Laboratory (KJ-14-109) and the Fundamental Research Funds for the Central Universities (JB161505).

Compliance with ethical standards

Conflict of interest Li Yang, Yafeng Qiao, Zhihong Liu, Jianfeng Ma and Xinghua Li declare that there are no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Bader DA, Kintali S, Madduri K, Mihail M (2007) Approximating betweenness centrality. In: Bonato A, Chung FRK (eds) Algorithms and models for the web-graph. Springer, San Diego, pp 124–137

Bakshy E, Karrer B, Adamic LA (2009) Social influence and the diffusion of user-created content. In: Proceedings of the 10th ACM conference on electronic commerce, ACM, pp 325–334

Bonacich P (2007) Some unique properties of eigenvector centrality. Social Netw 29(4):555–564

Brandes U (2001) A faster algorithm for betweenness centrality*. J Math Sociol 25(2):163–177

Brandes U (2008) On variants of shortest-path betweenness centrality and their generic computation. Social Netw 30(2):136–145

Catanese S, Ferrara E, Fiumara G (2013) Forensic analysis of phone call networks. Social Netw Anal Min 3(1):15–33

Domingos P, Richardson M (2001) Mining the network value of customers. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 57–66

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40(1):35–41

Freeman LC (1978) Centrality in social networks conceptual clarification. Social Netw 1(3):215–239

Fu Z, Sun X, Linge N, Zhou L (2014) Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonym query. IEEE Trans Consum Electron 60(1):164–172

Fu Z, Ren K, Shu J, Sun X (2016) Enabling personalized search over encrypted outsourced data with efficiency improvement. IEEE Trans Parallel Distrib Syst 27(9):2546–2559

Granovetter MS (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380

Guo P, Wang J, Geng XH, Kim CS, Kim JU (2014) A variable threshold-value authentication architecture for wireless mesh networks. J Internet Technol 15(6):929–935

Holme P, Kim BJ, Yoon CN, Han SK (2002) Attack vulnerability of complex networks. Phys Rev E 65(5):056,109

Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining, ACM, pp 137–146

Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. Nat Phys 6(11):888–893

Li J, Chen X, Li M, Li J, Lee PPC, Lou W (2013) Secure deduplication with efficient and reliable convergent key management. IEEE Trans Parallel Distrib Syst 25(6):1615–1625

Li J, Huang X, Li J, Chen X (2014) Securely outsourcing attribute-based encryption with checkability. IEEE Trans Parallel Distrib Syst 25(8):2201–2210

Li J, Li J, Chen X, Jia C, Lou W (2015) Identity-based encryption with outsourced revocation in cloud computing. IEEE Trans Comput 64(2):425–437

Lu Z, Fan L, Wu W, Thuraisingham B, Yang K (2014) Efficient influence spread estimation for influence maximization under the linear threshold model. Comput Soc Netw 1(1):1–19

Newman ME, Watts DJ (1999) Renormalization group analysis of the small-world network model. Phys Lett A 263(4):341–346

Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási AL (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci 104(18):7332–7336

Putzke J, Takeda H (2016) Identifying key opinion leaders in evolving co-authorship networksa descriptive study of a proxy variable for betweenness centrality. In: Cherifi H, Gonçalves B, Menezes R, Sinatra R (eds) Complex networks, vol VII. Springer, Switzerland, pp 311–323

Qin Y, Ma J, Gao S (2016) Efficient influence maximization under TSCM: a suitable diffusion model in online social networks. Soft Comput 1–12. doi:10.1007/s00500-016-2068-3

Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. In: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining, ACM, pp 61–70

Sabidussi G (1966) The centrality index of a graph. Psychometrika 31(4):581–603

Said YH, Wegman EJ, Sharabati WK, Rigsby JT (2008) Retracted: social networks of author-coauthor relationships. Comput Stat Data Anal 52(4):2177–2184

Saito K, Nakano R, Kimura M (2008) Prediction of information diffusion probabilities for independent cascade model. In: Lovrek I, Howlett RJ, Jain LC (eds) Knowledge-based intelligent information and engineering systems. Springer, Croatia, pp 67–75

Segarra S, Ribeiro A (2014) A stable betweenness centrality measure in networks. In: 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 3859–3863

Shen J, Tan H, Wang J, Wang J, Lee S (2015) A novel routing protocol providing good transmission reliability in underwater sensor networks. J Internet Technol 16(1):171–178

Wang M, Ma J (2015) A novel recommendation approach based on users weighted trust relations and the rating similarities. Soft Comput 1–10. doi:10.1007/s00500-015-1734-1

Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. Nature 393(6684):440–442

Xia Z, Wang X, Sun X, Wang Q (2015) A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data. IEEE Trans Parallel Distrib Syst 27(2):1–1

Zhang Z, Wang K (2013) A trust model for multimedia social networks. Social Netw Anal Min 3(4):969–979