

A new opinion leaders detecting algorithm in multi-relationship online social networks

Gengxin Sun¹  · Sheng Bin¹

Received: 19 January 2017 / Revised: 7 April 2017 / Accepted: 26 April 2017 /

Published online: 29 November 2017

© Springer Science+Business Media, LLC, part of Springer Nature 2017

Abstract Opinion leaders in online social networks are important for various fields such as public opinion propagation, marketing management, administrative science and even politics. There are often many kinds of relationships in an online social network. Detecting and identifying opinion leaders depending on any one kind of relationship is inaccurate. In this paper, node importance analysis in multi-relationship online social networks was proposed by signalling based on Multi-subnet Compositated Complex Networks Model, and considering the characteristics of multiple relationships which would interrelate with each other. Through node importance under multiple relationships, the novel opinion leader detecting algorithm in multi-relationship online social networks is proposed and approved to be efficient by experiments described in this paper.

Keywords Opinion leaders · Complex network · Signalling · Node importance · Online social network

1 Introduction

The concept of opinion leaders was first introduced by Lazarsfeld [10], who is a sociologist, for a group of individuals who have the power to affect a group's perceptions. The concept of opinion leaders was developed by Von Hippel [21, 22] to characterize a set of users that are ahead of others in terms of developing new product needs, and was subsequently taken up by other researchers in social networks, aiming to understand users whose opinions are highly influenced by these opinion leaders.

The concept of opinion leaders is widely used in online social network analysis as they emerge in the interaction between subjects. There have been many researchers from different

✉ Gengxin Sun
sungengxin@qdu.edu.cn

¹ College of Data Science and Software Engineering, Qingdao University, Qingdao 266071, People's Republic of China

aspects to research opinion leaders detection. Ravi [17] presents a rigorous survey on opinion mining and sentiment analysis, which portrays views presented by over one hundred articles published in the last decade regarding necessary tasks, approaches, and applications. So far, opinion leaders are being detected and identified using standard questionnaires or observed behaviour. Belz [1] has shown that using the analysis of posts in internet forums, classifies almost half of internet users correctly as opinion leaders or non-leaders. Bilgram [2] studied online social network communities and identified several factors for identification of opinion leaders. Mazhari [11] proposed a model for measuring the similarities of two members based on information contained in their profiles in online social networks. The model can contribute to detect opinion leaders through friendship relationship. Severyn [18] designed the model for effectively carrying out opinion mining on YouTube comments. Huang [7] used the comments in online social networks to detect the opinion leaders. It can find the most influential comments from all the network comments and the most influential users form the entire user network. Above studies mainly depend on comments or sentiment analysis to detect opinion leaders and require comprehensive analysis of opinions expressed in online social networks. But analysis of opinions expressed is very difficult and the most influential comments may not belong to the most influential opinion leaders.

On the other hand, Kratzer [9] detected opinion leaders by specific network topologies, specifically some typical topologies of complex network, such as betweenness and degree centrality. Online social networks are typical examples of complex networks and complex network theories can be used to analyse their network topological characteristics. There are many traditional evaluation indicators for node importance in complex networks. These indicators are mainly divided into two categories: local connectivity importance indicators of node, and global connectivity importance indicators of node. Degree and clustering coefficient are typical indicators of the former, betweenness, closeness and composite indicator for node importance are typical indicators of the latter.

The degree of a node in a network is the number of connections or links that the node has to other nodes. High-degree nodes are those that have plenty of nodes connected with them. It is revealed that high-degree nodes are more important in a manner [3, 6, 20, 23]. The indicator is very intuitive and has low time complexity, but it can only reflect the local importance of a node.

Betweenness of a node is the total number of shortest paths from all nodes to all others that pass through that node. This indicator can reflect perfectly a node's centrality in a network [5, 12]. A node with high betweenness has a large influence on the transfer of items through the network. The dynamic change characteristics of betweenness over a certain timeframe are taken into consideration. Dynamic betweenness [8, 15], such as flow betweenness and random-walk betweenness are proposed.

The farness of a node is defined as the sum of its shortest paths to all other nodes, and its closeness is defined as the reciprocal of the farness [14]. The indicator can reflect the impact of the node on other nodes, as well as the impact of topology location for nodes.

In online social networks, opinion leaders can be defined as the impact of node importance in complex networks. But there are many social relationships, such as friendship, commerce, or others in online social networks. With regards to opinion leaders, they can have different interests for different subjects, so groups of opinion leaders tend to overlap. It is inaccurate to detect and identify opinion leaders in multi-relationship online social networks according to a single kind of relationship alone. The aim of our study is to fill those gaps. Differing from previous studies [1, 2, 4, 7, 9, 11, 13, 18, 19, 24, 25], we propose a complex network model,

multi-subnet composited complex network, in which a link represents all kinds of relationships between nodes. Given comments or sentiment analysis similarity relationships and network topologies, subnets could be picked up from the original network. Through multi-subnet composited networks, multiple relationships in online social networks can be used to detect and identify opinion leaders accurately.

2 Multi-subnet composited complex network model

Definitions 1 (Multi-subnet Composited Complex Network): A multi-subnet composited complex network is 4-tuple $G = (V, E, R, F)$, where

- (1) $V = \{v_1, v_2, \dots, v_m\}$ is a finite set of nodes and $m = |V|$;
- (2) $E = \{\langle v_h, v_l \rangle \mid v_h, v_l \in V, 1 \leq h, l \leq m\} \subseteq V \times V$ is a finite set of links between nodes;
- (3) $R = R_1 \times \dots \times R_i \times \dots \times R_n = \{(r_1, \dots, r_i, \dots, r_n) \mid r_i \in R_i, 1 \leq i \leq n\}$, where R_i is one kind of relationships between nodes and the amount of kinds of relationships is n ;
- (4) F is a mapping from E to R .

Mapping F gives each link an n -tuple to denote all kinds of relationships between its two nodes. Let $r_i = \emptyset$ be no interrelation R_i between the nodes.

Definitions 2 (Subnet): Assuming multi-subnet composited complex network $G = (V, E, R, F)$, $G' = (V', E', R', F')$ is said to be subset of G according to set of relationships R' ($R' \subseteq R$ and $R' \neq \emptyset$), iff

- (1) $V' \subseteq V$;
- (2) $E' = \{\langle v_h, v_l \rangle \mid F(\langle v_h, v_l \rangle) \cap R' \neq \emptyset, \langle v_h, v_l \rangle \in E, v_h, v_l \in V\}$;
- (3) $F' : E' \rightarrow R'$, where $\forall \langle v_h, v_l \rangle \in E', \text{ s.t. } F'(\langle v_h, v_l \rangle) = F(\langle v_h, v_l \rangle)$.

Actually, a subnet is also a composited network. The reason that we define it is to show a composited network could be dynamically organized. Due to the complexity of complex systems, we usually model one aspect of these systems as a subnet in which one kind of relationships among parts of the system is considered. Whereas for some problems, comprehensive analysis referring to several kinds of relationships in the system is essential. In that case, we expect that several subnets could be composited together.

For practical usage, multi-dimensional space is imported into our model, converting the relationships into a spatial vector to make it computable for solving real problem. Based on multi-subnet composited complex network, a vector composited network is defined by converting the relationships into a spatial vector.

Just as each link is assigned a weight proportional to the intensity or capacity of multiple relationships in a weighted network, $dom(r_i)$ is defined to denote the intensity set. Each element in $dom(r_i)$ indicates the intensity of relationship r_i .

Definitions 3 (Vector Space of relationship intensity): Assuming a multi-subnet composited complex network $G = (V, E, R, F)$, where $R = R_1 \times \dots \times R_i \times \dots \times R_n = \{(r_1, \dots, r_i, \dots, r_n) \mid r_i \in R_i, 1 \leq i \leq n\}$, let $dom(r_i) \subseteq \mathcal{R}^+$, null intensity is zero.

Define

$$S = \text{dom}(r_1) \times \dots \times \text{dom}(r_i) \times \dots \times \text{dom}(r_n) \quad (1)$$

as set of ordered n -tuple real numbers. It is obvious that S is vector-space of a field of real numbers. S is a vector-space of relationship intensity. Eq.(1) could be rewrote as Eq.(2) below,

$$S = \{s = (s(r_1), \dots, s(r_i), \dots, s(r_n)) | s(r_i) \in \text{dom}(r_i), 1 \leq i \leq n\} \quad (2)$$

where s is the vector of relationship intensity and $s(r_i)$ is the i -th term of s denoting the intensity of relationship r_i .

Definitions 4 (Bases of Vector Space of interrelation intensity): Assuming multi-subnet composited complex network $G = (V, E, R, F)$, where $R = R_1 \times \dots \times R_i \times \dots \times R_n = \{(r_1, \dots, r_i, \dots, r_n) | r_i \in R_i, 1 \leq i \leq n\}$. Let S be a vector space of relationship intensity G , set of vectors

$$e_1 = (1, 0, \dots, 0), e_2 = (0, 1, \dots, 0), \dots, e_n = (0, 0, \dots, 1) \quad (3)$$

is defined as bases of S . For $\forall s \in S$, $s = s(r_1)e_1 + s(r_2)e_2 + \dots + s(r_n)e_n$. $e_1, \dots, e_i, \dots, e_n$ are linearly independent vectors. Base e_i could be regarded as relationship r_i and vice versa. The number of dimensions of S is $\text{dim}(S) = n$.

Definitions 5 (Vector Network): Given $G = (V, E, R, F)$, let S be a vector space of relationship intensity G . Then there is the mapping $M: E \rightarrow S$, s.t. for $\forall \langle v_h, v_l \rangle \in E$, $v_h, v_l \in V, 1 \leq h, l \leq m$

$$M(\langle v_h, v_l \rangle) = s^{\langle v_h, v_l \rangle} = (s^{\langle v_h, v_l \rangle}(r_1), \dots, s^{\langle v_h, v_l \rangle}(r_i), \dots, s^{\langle v_h, v_l \rangle}(r_n)) \quad (4)$$

and if $r_i \neq \emptyset$ in $F(\langle v_h, v_l \rangle)$ then $s^{\langle v_h, v_l \rangle}(r_i) \neq 0, 1 \leq i \leq n$. Three-tuple $\Sigma = (G, S, M)$ is a vector network of G . And let $s^{\langle v_h, v_l \rangle}(r_i)$ indicates the intensity of relationship r_i between nodes v_h, v_l .

3 Node importance evaluation approach in composited networks based on signalling

3.1 Signalling process

For a network with nodes, every node is assumed to be a system which can send, receive and record signals. One node can only affect its neighbour nodes, likewise those nodes also will affect their neighbour nodes in the same way. With the signalling process, each node will affect the whole network.

At the beginning, a node as source is selected and has one unit of signal and the other nodes in the network have no signal. Then the source node sends a signal to all of its neighbour nodes and itself. After the first process, the node and all its neighbour nodes have a signal. All the nodes which have a signal will subsequently send it to their neighbour nodes and themselves. Every node would record the quantity of signals which it received and then would send the same quantity of signals in the next process. In this way, the signalling process would be

repeated constantly on the network. After T time process, an nd vector that records each node's signal quantity which represents the impact of the source node is received. The signalling process of network with 6 nodes is shown in Fig. 1a-d.

The signalling process in Fig. 2 could be described by using a mathematical equation shown as follows:

$$V = (I + A)^T \tag{5}$$

where, A is an adjacency matrix, which can represent mathematically a network with s nodes, elements A_{ij} of the adjacency matrix represents whether there is a link from node i to node j and 0 otherwise. I is an identity matrix. T represents the signalling process time. So the column i of matrix V represents the impact of source node i to the whole network in signalling process time, T .

In order to get the relative impact, each row of matrix V should be standardized. Assume the row i of V is $V_i = (v_{i1}, v_{i2}, \dots, v_{is})$, then the V_i can be standardized as $U_i = (u_{i1}, u_{i2}, \dots, u_{is})$, here $u_{ij} = v_{ij} / \sum_j v_{ij}$.

3.2 Signalling process in multi-subnet composited complex networks

Every element \hat{A}_{ij} of adjacency matrix \hat{A} of a multi-subnet composited complex network is an m -tuple which represents all kinds of relationships of link from node i to node j . The i th vector component $s^{<v_i, v_j>}(r_f)$ represents if there is relationship r_f on link from node i to node j or not.

In a multi-subnet composited complex network, when the source node sends signal about relationship r_f to all of its neighbour nodes and itself, only those neighbour nodes which are connected to the source node to form a link where there is relationship r_f can receive signals. Other neighbour nodes connected with the source node to form link where there is no relationship r_f cannot receive signals with regard to relationship r_f .

After T signalling process, for every source node, an nd vector that includes all relationships in the network occurs, every vector component is all m -tuple $(r_1, \dots, r_i, \dots, r_m)$ which represents all kinds of relationships in the network. Thus, for the whole multi-subnet composited complex network, an $n * n$ adjacency matrix whose every element is a m -tuple occurs.

The signalling process of a multi-subnet composited complex network with 6 nodes and node 1 acting as source node is shown in Fig. 2a-d.

The signalling process in a multi-subnet composited complex network could be described by using a mathematical equation shown as follows:

$$\hat{V} = (\hat{I} + \hat{A})^T \tag{6}$$

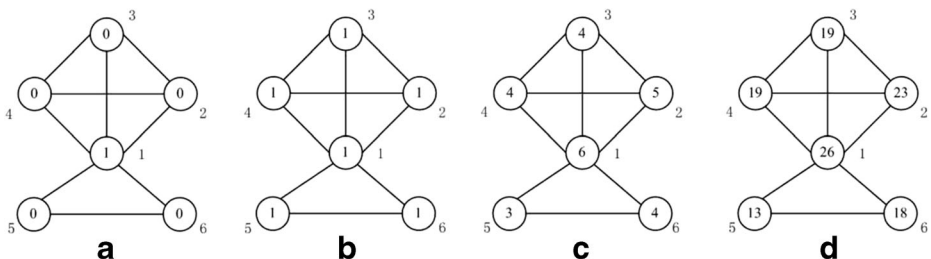


Fig. 1 Signalling process in network with 6 nodes

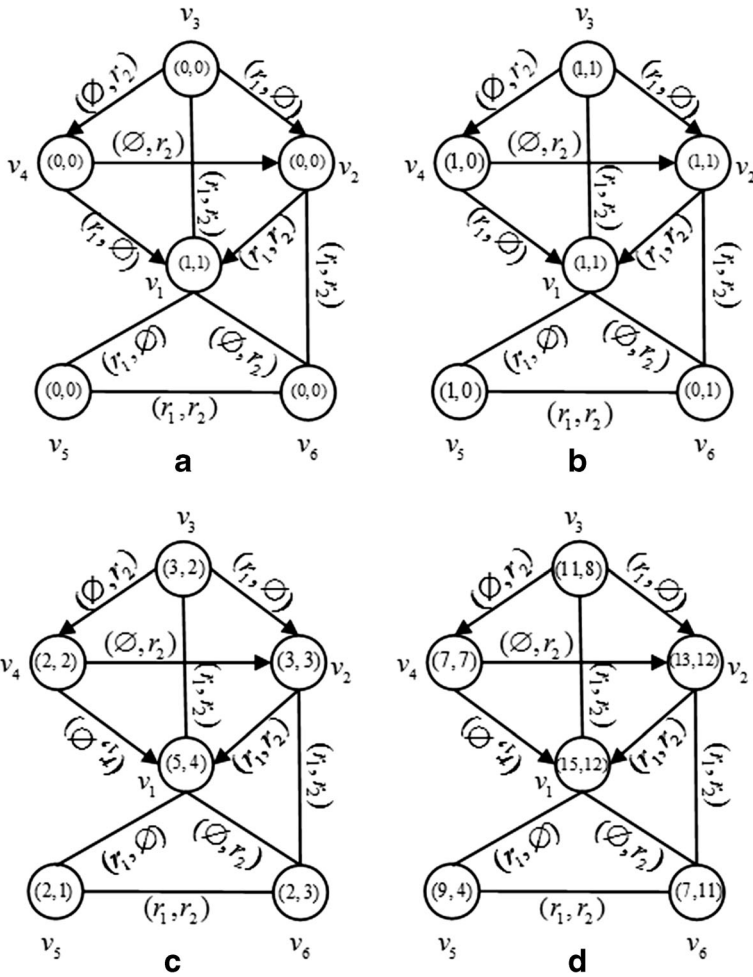


Fig. 2 Signalling process of multi-subnet composed complex network with 6 nodes

where, T represents the signalling process time. \hat{A} is an adjacency matrix of a multi-subnet composed complex network, it can be expressed as follows:

$$\hat{A} = \begin{matrix} & v_1 & v_2 & v_3 & v_4 \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{matrix} & \begin{bmatrix} (0,0) & (1,1) & (1,1) & (0,1) \\ (1,1) & (0,0) & (1,1) & (0,0) \\ (1,1) & (1,1) & (0,0) & (0,0) \\ (0,1) & (0,0) & (0,0) & (0,0) \end{bmatrix} \end{matrix} \tag{7}$$

The above signalling process is under the assumption that there is no interrelation between the multiple relationships. Yet in reality, there would be interrelation between the multiple relationships in a multi-subnet composed complex network. For a multi-subnet composed complex network, the parameter α can be used to represent the interrelationship between multiple relationships in the signalling process.

For example, if relationship r_2 effects relationship r_1 , signalling t_0 on relationship r_1 which is received by neighbour nodes v_0 of source node v_s , can be expressed as follows:

$$t_o = \begin{cases} \alpha \times 1 & \text{if } s^{<v_s, v_o>}(r_1) = 1 \text{ and } s^{<v_s, v_o>}(r_2) = 0 \\ 1 & \text{if } s^{<v_s, v_o>}(r_1) = 1 \text{ and } s^{<v_s, v_o>}(r_2) = 1 \end{cases} \quad (8)$$

The signalling process of an interactional multi-subnet composited complex network with 6 nodes and node 1 acting as source node is shown in Fig. 3a-d.

4 Opinion leader detecting algorithm in multi-relationship online social networks

The signalling process in multi-relationship online social networks could be described by a matrix iterative operation based on a multi-subnet composited complex network. For node v_i in

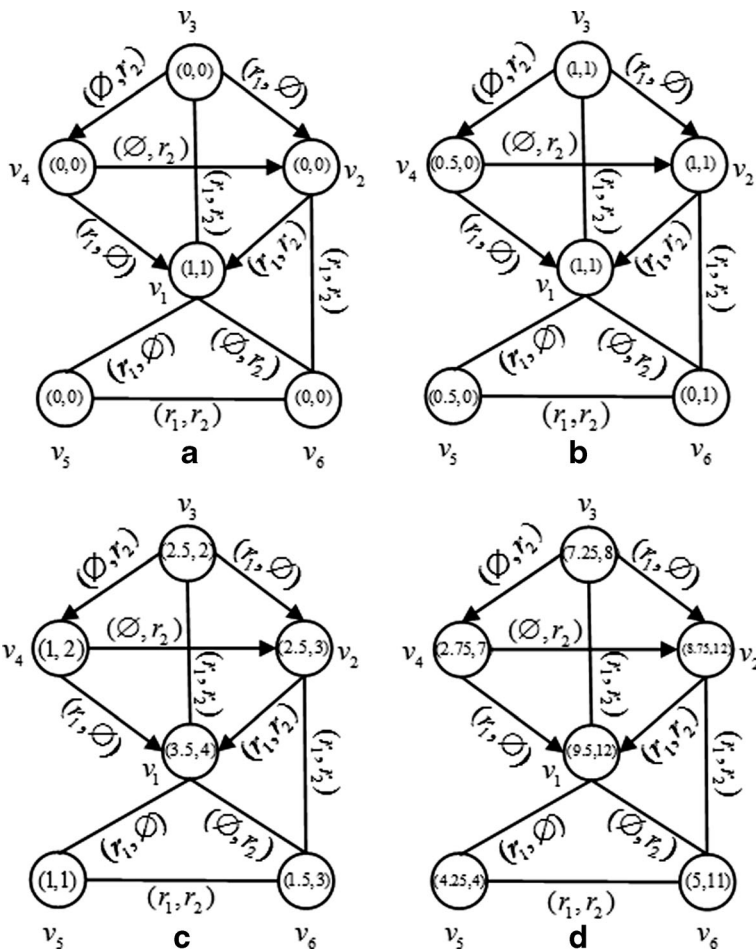


Fig. 3 Signalling process of interactional multi-subnet composited complex network with 6 nodes. In the process, parameter $\alpha = 0.5$

a multi-subnet composited complex network $G = (V, E, R, F)$, $\hat{k}_{v_h}^{r_i}$ represents the degree of node v_h on relationship r_j , expressed as follows:

$$\hat{k}_{v_h}^{r_i} = \sum_{v_l \in V} \hat{a}_{hl}(r_i) \tag{9}$$

The node degree on relationship r_j indicates the number of nodes on which signalling on relationship r_j can be directly reached. Importance vector R can be conducted by an adjacency matrix, and importance matrix D can be constructed by an adjacency matrix and importance vector R . The steps of an opinion leader detecting algorithm are described as follows:

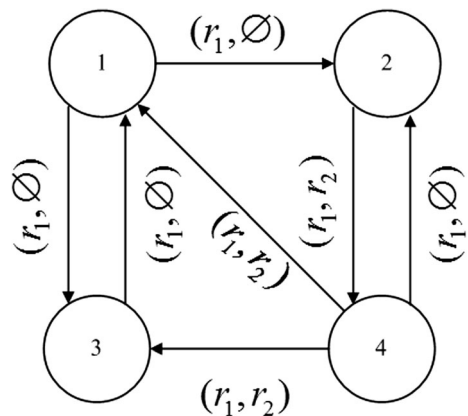
- (1) Generating the initial adjacency matrix \hat{A} of a multi-relationship online social network;
- (2) Generating the initial importance vector R of every node by adjacency matrix \hat{A} ;
- (3) Constructing an importance matrix D by importance vector R and adjacency matrix \hat{A} ;
- (4) Using matrix multiplication to simulate the signalling process, the matrix multiplication mathematical equation is defined as follows: $R_{i+1} = R_i \cdot D_i$;
- (5) Comparing R_{i+1} with R_i to judge that the difference is whether or not more than setting threshold value, which means that importance vector R is whether or not tending towards stability after matrix iterative, if stable state is achieved, the algorithm would terminate, otherwise, jumping to step (3) for going on matrix iterative operation.

5 Experiment and analysis

To verify our proposed algorithm, experimental analysis is conducted. The data for the experiment consists of 30,453 users in Douban.com [16] through tracking their relationships and comments over two days (2015–03-17 08:00: 00~2015–03-18 07:59:59). The experimental network is composed of 30,453 nodes, 55,865 links and 1,789,567 book comments. There are friendship relationships and book comments similar in relationship in the network, and the two kinds of relationships interrelate with each other. So the experimental network is a typical multi-relationship online social network.

Firstly, we select 4 users from the data to form a directed network to validate node importance of our algorithm. The 4 users network are shown as Fig. 4.

Fig. 4 The 4 users network used to validate our algorithm



We assumed that if the difference in every node importance in two iterations is less than 10^{-3} , the important degree of the nodes is regarded as stable, and then the termination condition of the algorithm would be met.

Comparison of the importance degree of nodes of our algorithm with the PageRank algorithm and HITS algorithm for a 4 user network is shown in Table 1.

From Table 1 we can see that if the multiple relationships in a 4 user network are ignored, the analysis results of the nodes importance of a 4 user network through our algorithm and classical algorithms, such as the PageRank algorithm and HITS algorithm, are completely consistent.

If the difference of every node importance in two iterations is less than a certain threshold, the important degree of nodes is regarded as stable. The iteration times with different thresholds to display trend stability through our algorithm, the PageRank algorithm and HITS algorithm are shown in Fig. 5.

From Fig. 5 we can see that the convergence of our proposed algorithm is much better than the PageRank and HITS algorithms. In the case of the same importance difference threshold, the iteration time of our proposed algorithm is much less than the iteration times of the PageRank and HITS algorithms.

In the experiment, First of all, manual voting method is used to detect opinion leaders. The first 500 users got by HITS algorithm and PageRank algorithm were investigated with email anonymous questionnaire, and 500 candidate opinion leaders were given in the questionnaire. A total of 378 valid questionnaires were returned, according to the results of the questionnaire, the top 20 opinion leaders were obtained.

Then through friendship relationship building of a single relationship network, the PageRank and HITS algorithms are used to detect opinion leaders, and via multiple relationships in the same data building a composited network, applying the algorithm proposed in Section 4, the top 20 nodes in order of importance are shown in Table 2.

From Table 2 we can see that though the sequence of the top 20 nodes is different for each of the three algorithms, there are only 3 nodes in the top 20 node sequences of our proposed algorithm not in the sequence of the PageRank and HITS algorithms. If the top 100 and top 500 nodes of the three algorithms sequences are selected, the three sequences are completely covered. So for node importance analysis, the three algorithms can be considered to be concordant.

But the detecting important nodes of our proposed algorithm are very close to the results of manual voting, and there are very big difference between the other two algorithms and the results of manual voting. It is proved that our proposed algorithm considering multiple relationships joint effects are more in line with the reality.

Table 1 Comparison of important degree of nodes of our algorithm with PageRank algorithm and HITS algorithm

	Our proposed algorithm	PageRank algorithm	HITS algorithm
Iteration times	11	25	16
Important degree of node 1	2.243	1.231	1.000
Important degree of node 2	1.586	0.923	0.732
Important degree of node 3	1.586	0.923	0.732
Important degree of node 4	1.586	0.923	0.732
Important degree average of nodes	1.750	1.000	0.800

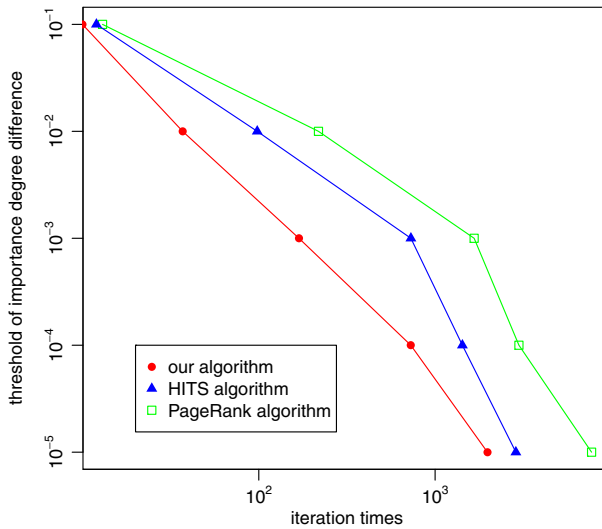


Fig. 5 Iterative times comparison of our algorithm PageRank algorithm and HITS algorithm

For the purpose of validating our algorithm further, sentiment analysis of comments is applied, where comments are usually divided into positive (1), negative (-1) and neutral (0). We divide the period of tracking data time into two different time periods. There are 211 comments of top 500 importance nodes in the first time period as shown in Fig. 6.

Table 2 Comparison of the analysis results of our proposed algorithm, PageRank algorithm and HITS algorithm with manual voting

	Manual voting	Our proposed algorithm	PageRank algorithm	HITS algorithm
1	No.701	No.701	No.6939	No.6939
2	No.7018	No.7018	No.3303	No.3303
3	No.174	No.1239	No.701	No.701
4	No.3377	No.3356	No.8220	No.8220
5	No.3356	No.174	No.7018	No.7018
6	No.7132	No.209	No.1239	No.3549
7	No.4589	No.3549	No.3549	No.1239
8	No.209	No.7132	No.286	No.286
9	No.2828	No.3561	No.3356	No.3356
10	No.6461	No.4323	No.7911	No.2914
11	No.3303	No.3303	No.174	No.174
12	No.3549	No.702	No.13237	No.13237
13	No.6939	No.2828	No.286	No.286
14	No.4323	No.6461	No.702	No.3561
15	No.8220	No.4589	No.209	No.209
16	No.702	No.6939	No.7132	No.702
17	No.4113	No.2914	No.3561	No.2828
18	No.377	No.8220	No.2828	No.7132
19	No.7911	No.13237	No.2914	No.7911
20	No.13237	No.7911	No.6395	No.6395

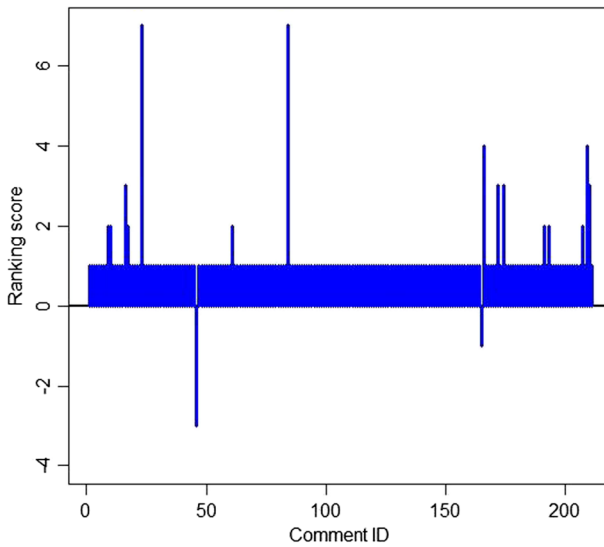


Fig. 6 Ranking (08:00:00–12:30:00)

It is also easy to find that comment No.25 which belongs to user No.1239 received the highest scores, while comment No.166 which belongs to user No.4323 didn't receive the highest score. Hence, user No.4323 is not an opinion leader at this time. As illustrated in Fig. 7, user No.4323 should be an opinion leader, while the score of user No.166 increases with time. But opinion leader user No.4323 can't be detected by either the PageRank or HITS algorithms.

Consequently, it is found that our algorithm proposed in this paper has much more accuracy than other algorithms, and our algorithm can better apply to detecting opinion leaders of multi-relationship online social networks.

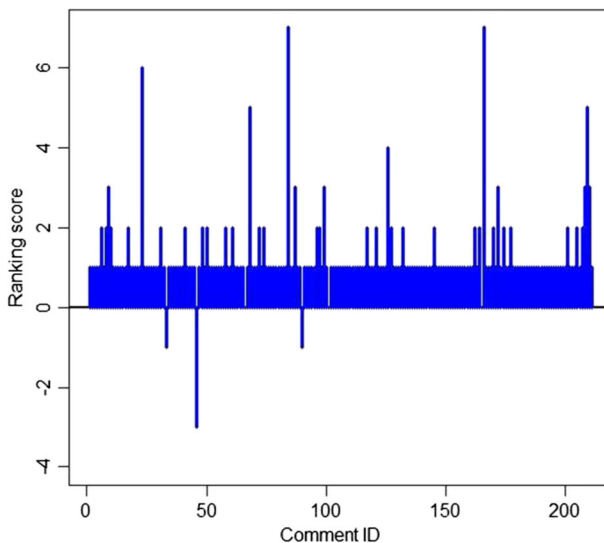


Fig. 7 Ranking (08:00:00–next day 07:59:59)

6 Conclusion

Based on a multi-subnet composited complex network model and signalling process, we proposed the use of an importance matrix iterative method to analyse node importance, and thereby detect opinion leaders in multi-relationship online social networks. Through experiment analysis, it proved that compared with existing node importance evaluation approaches, our proposed algorithm can better reflect node importance in multi-relationship online social networks, and iteration times of our proposed algorithm are much lower. The research would contribute to improved efficiency and accuracy of public opinion analysis, community structure detecting, and signalling processing in multi-relationship online social networks.

Acknowledgements Many thanks are due to Neal Gilmore and Karen Gilmore for their assistance in language revision. Also, thanks to the editor and all anonymous reviewers for their constructive comments.

Funding This work is supported by the Humanity and Social Science Youth foundation of Ministry of Education of China (grant no. 15YJC860001). This research is also supported by the Social Science Foundation of Qingdao, China (grant no. QDSKL150437), Statistical Science Research Project of China (grant no. 2015LZ20) and China Postdoctoral Science Foundation Funded Project (grant no. 2015 M580571, grant no. 2016 M590612).

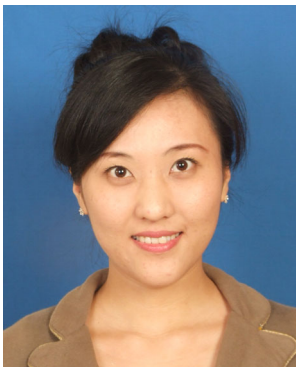
References

1. Belz FM, Baumbach W (2010) Netnography as a method of lead user identification. *Creat Innov Manag* 19(3):304–313
2. Bilgram V, Brem A, Voigt KI (2008) User-centric innovations in new product development—systematic identification of lead users harnessing interactive and collaborative online-tools. *Int J Innov Manag* 12(03):419–458
3. Callaway D, Newman J, Strogatz S (2000) Network robustness and fragility: percolation on random graphs. *Phys Rev Lett* 85:5468–5471
4. Ding, C., He, X., Husbands, P., Zha, H., and Simon, H. D. PageRank, HITS and a unified framework for link analysis. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, 2002, pp. 353–354
5. Freeman L (1977) A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41
6. Haibo H, Xiaofan W (2008) Analysis of online social networks based on complex network theory. *Complex Syst Complexity Sci* 2(5):1–14
7. Huang B, Yu G and Karimi H R. The finding and dynamic detection of opinion leaders in social network. *Probl Eng* 2014
8. Kim H, Tang J, Aderson R (2012) Centrality prediction in dynamic human contact networks. *Comput Netw* 56(3):983–996
9. Kratzer J, Lettl C (2009) Distinctive roles of lead users and opinion leaders in the social networks of schoolchildren. *J Consum Res* 36(4):646–659
10. Lazarsfeld PF, Berelson B, Gaudet H (1965) *The people's choice. How the voter makes up his mind in a presidential campaign.* Columbia University Press, New York
11. Mazhari S, Fakhrahmad SM, Sadeghbeygi H (2015) A user-profile-based friendship recommendation solution in social networks. *J Inf Sci* 41(3):284–295
12. Milgram S (1967) The small world problem. *Psychology today* 2(1):60–67
13. Miller, J. C., Rae, G., Schaefer, F., Ward, L. A., LoFaro, T., and Farahat, A. Modifications of Kleinberg's HITS algorithm using matrix exponentiation and web log records. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, 2001, pp. 444–445
14. Okamoto K, Chen W, Li X (2005) Ranking of closeness centrality for large-scale social networks. *Lect Notes Comput Sci* 59:186–195
15. Opsahl T, Agneessens F, Skvoretz J (2011) Node centrality in weighted networks: generalizing degree and shortest paths. *Soc Networks* 32(1):245–251
16. Qingqing Z (2013) Relationship between scores and tags for Chinese books-in the case of Douban book. *Chinese J Libr Inf Sci* 6(4):40–54
17. Ravi K, Ravi VA (2015) Survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowl-Based Syst* 89:14–46

18. Severyn A, Moschitti A, Uryupina O (2015) Multi-lingual opinion mining on YouTube. *Inf Process Manag* 52(1):46–60
19. Song, X, Chi, Y, Hino, K, and Tseng, B. Identifying opinion leaders in the blogosphere. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, 2007, pp. 971–974
20. Song X, Yan X, Li Y (2015) Modelling liking networks in an online healthcare community: an exponential random graph model analysis approach. *J Inf Sci* 41(1):89–96
21. Von Hippel E (1986) Lead users: a source of novel product concepts. *Manag Sci* 32(7):791–805
22. Von Hippel E (2007) The sources of innovation. *Das Summa Summarum des Management*. Gabler, p 111–120
23. Xia Y and Fan J. Efficient attack strategy to communication networks with partial degree information. 2011 I.E. Int Symp, 2011, pp.1091–1095
24. Yu X, Wei X, Lin X (2010) Algorithms of BBS opinion leader mining based on sentiment analysis. *Web Information Systems and Mining*. Springer Berlin Heidelberg, Heidelberg, pp 360–369
25. Zhai Z, Xu H, Jia P (2008) Identifying opinion leaders in BBS. *Web Intelligence Intell Agent Technol*:398–401



Gengxin Sun received his Ph.D. degree in Computer Science from Qingdao University, China in 2013. He is currently an Associate Professor in the School of Computer Science and Engineering at Qingdao University. His main research interests include embedded system, operating system, complex networks, web information retrieval and data mining.



Sheng Bin received her Ph.D. degree in Computer Science from Shandong University of Science and Technology, China in 2009. She is currently a lecturer in the School of Software Technology at Qingdao University, China. Her main research interests include embedded system, operating system, complex networks, cloud computing and data mining.