International Journal of Hybrid Information Technology
Vol. 9, No. 5 (2016), pp. 391-398
http://dx.doi.org/10.14257/ijhit.2016.9.5.33

An Opinion Leaders Detecting Algorithm in Multi-relationship Online Social Networks

[1]Weihua Zhang, [1]Gengxin Sun and [2]Sheng Bin

[1] International College of Qingdao University, Qingdao, China
[2]Software Technical College of Qingdao University, Qingdao, China
zhangweihua@qdu.edu.cn

Abstract

Opinion leaders in online social networks are important for a lot of applications in 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 all accurate. In this paper, node importance analysis in multi-relationship online social network was proposed based on signaling spreading, and considering the characteristics of multiple relationships which would interrelate with each other. Through node importance, a novel opinion leaders detecting algorithm is proposed and approved to be efficient by experiments described in the paper.

Keywords*: opinion leader, complex network, signaling spreading, node importance, online social network*

1. Introduction

The concept of opinion leader was introduced firstly by Lazarsfeld [1], who is a sociologist, for a group of individuals who have the power to affect a group's perceptions. The concept of opinion leader was developed by Von Hippel [2-3] to characterize a set of users that are ahead to others in terms of developing new product needs. The concept of opinion leader was subsequently taken up also by researchers in social networks, aiming to understand users whose opinions are highly influenced by those opinion leaders.

The concept of opinion leader is widely used in online social network analysis as they emerge in the interaction between subjects. So far, opinion leaders are being detected and identified using standard questionnaires or observed behavior. Belz [4] have shown that using the analysis of posts in internet forums, classifies almost half of internet users correctly as opinion leaders or non-leaders. Bilgram [5] studied online social network communities and identified several factors for identification of opinion leaders. These approaches all require comprehensive analysis of opinions expressed in online social networks. In contrast Kratzer [6] found out that opinion leaders by specific network topologies, specifically some typical topologies of complex network, such as betweenness and degree centrality for opinion leaders.

Online social networks are typical example of complex network, there are many social relationships, such as friendship, commerce, or others in the network. With regards to opinion leaders, they can have different interests for different subjects. The 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 kind of relationship alone.

ISSN: 1738-9968 IJHIT

2. Node Importance Evaluation Approaches of Complex Network

In this paper, we study of identifying opinion leaders in online social networks using only user interaction information through complex network theories. These interaction information determine node importance in network.

There are many measure indicators for node importance evaluation in complex networks. These indicators are mainly divided into two categories: one category is local connection importance indicators of node, another category is global connection 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 edges the node has to other nodes. High-degree nodes mean that there are plenty of nodes connected with them. It revealed that the high-degree nodes are more important in a manner [7-10]. The index is very intuitive and has low time complexity, but it can only reflect the local importance of node.

Betweenness of a node is the total number of shortest paths from all nodes to all others that pass through that node. The indicator can reflect perfectly a node's centrality in a network [11-12], a node with high betweenness has a large influence on the transfer of items through the network. The dynamic change characteristics of betweenness during a certain time are taken into consideration, dynamic betweenness [13-14], such as flow betweenness, 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 [15]. The indicator can reflect the impact of the node for other nodes, as well as the impact of topology location for nodes.

In online social networks, node importance can be defined as the impact of some node for information dissemination ability in networks. For above-mentioned traditional node importance evaluation indicators, degree only can represent information dissemination ability of nodes on the local scale. The indicators based on shortest paths, such as betweenness, closeness and so on only take nodes importance on topology location into consider, it would not reflect accurately the impact of nodes for information dissemination. Therefore, it is highly necessary to research importance nodes detecting algorithm which is suitable for its characteristics of online social networks.

3. Node Importance Evaluation Approach Based on Signaling Spreading

In complex networks, nodes can be converted into vectors of the algebraic space by using signaling spreading [16].

3.1. Signaling Spreading

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

At the beginning, a node as source is selected and let it has one unit of signal and the other nodes in network have no signal. Then the source node send signal to all of its neighbor nodes and itself. After the first spreading process, the node and all its neighbor nodes have a signal. All the nodes which have signal will subsequently send it to their neighbor 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 spreading process. In this way, the spreading process would be repeated constantly on the network. After T spreading process, a nd vector that records each node's signal quantity which represents the impact of the source node is got. The signaling spreading process of network with 6 nodes is shown in Fig. 1.

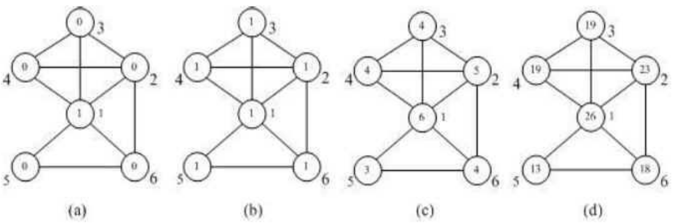

Figure 1. Signaling Spreading Process of Network with 6 Nodes

For the whole network with n nodes, n vectors will be got. The n vectors are projected to the algebraic space through standardization, every vector would be acted as data object in the algebraic space.

The above signaling spreading process could be described by using a mathematical equation shown as eq. (1).

$$V = (I + A)^T \tag{1}$$

Where, A is an adjacency matrix, which can represent mathematically a network with n nodes, elements A_{ij} of the adjacency matrix represents if there is an edge from node i to node j and 0 otherwise. I is identity matrix. T represents times of signaling spreading. So the column i of matrix V represents the impact of source node i to the whole network in T times of signaling spreading.

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

3.2. Signaling Spreading in Multi-Relationship Network

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

In multi-relationship network, when source node send signal about relationship r_f to all of its neighbor nodes and itself, only those neighbor nodes which connected with the source node to form edge on where there is relationship r_f can receive signals. Other neighbor nodes which connected with the source node to form edge on where there is no relationship r_f can receive signals with regard to relationship r_f.

In multi-relationship network, after T spreading process, for every source node, a nd vector that includes all relationships in network is got, every vector component is all m-tuple $(r_1, ..., r_i, ..., r_m)$ which represents all kinds of relationships in network. Thus, for the whole multi-relationship network, a $n*n$ adjacency matrix whose every element is a m-tuple is got.

The signaling spreading process of a multi-relationships network with 6 nodes and node 1 acting as source node is shown in Fig. 2.

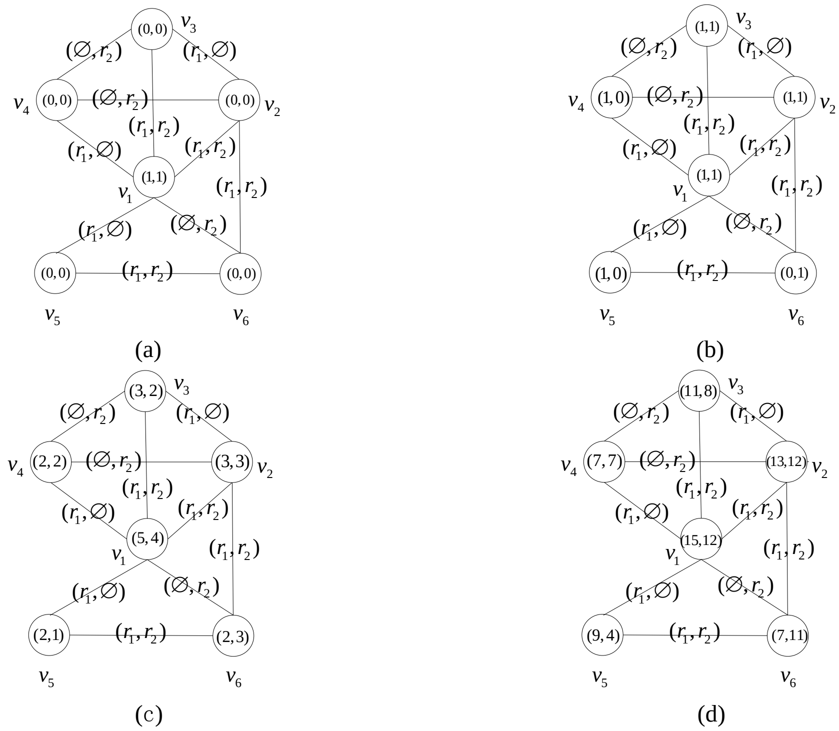

Figure 2. Signaling Spreading Process of Multi-Relationships Network with 6 Nodes

The signaling spreading process in multi-relationship network could be described by using a mathematical equation shown as eq. (2).

$$\hat{V} = (\hat{I} + \hat{A})^T \tag{2}$$

Where, T represents times of signaling spreading. $\hat{A}$ is an adjacency matrix of multi-relationship network, it can be expressed as follows:

$$\hat{A} = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{bmatrix} (0,0) & (1,1) & (1,1) & (0,1) \\ (1,1) & (0,0) & (1,1) & (0,0) \\ (1,1) & (1,1) & (0,0) & (0,0) \\ (0,1) & (0,0) & (0,0) & (0,0) \end{bmatrix} \end{array}$$

The above signaling spreading process is under the assumption that there is no interrelation between multiple relationships. Yet in reality, there would be interrelation between multiple relationships in multi-relationship online social networks. For such multi-relationship online social networks, parameter α can be used to represent interrelation between multiple relationships in signaling spreading process.

For example, if relationship r_2 effects relationship r_1, signaling t_o on relationship r_1 which is received by neighbor nodes v_o of source node v_s can be expressed as follows:

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

The signaling spreading process of an interactional multi-relationships network with 6 nodes and node 1 acting as source node is shown in Fig. 3.In the process, parameter $\alpha = 0.5$.

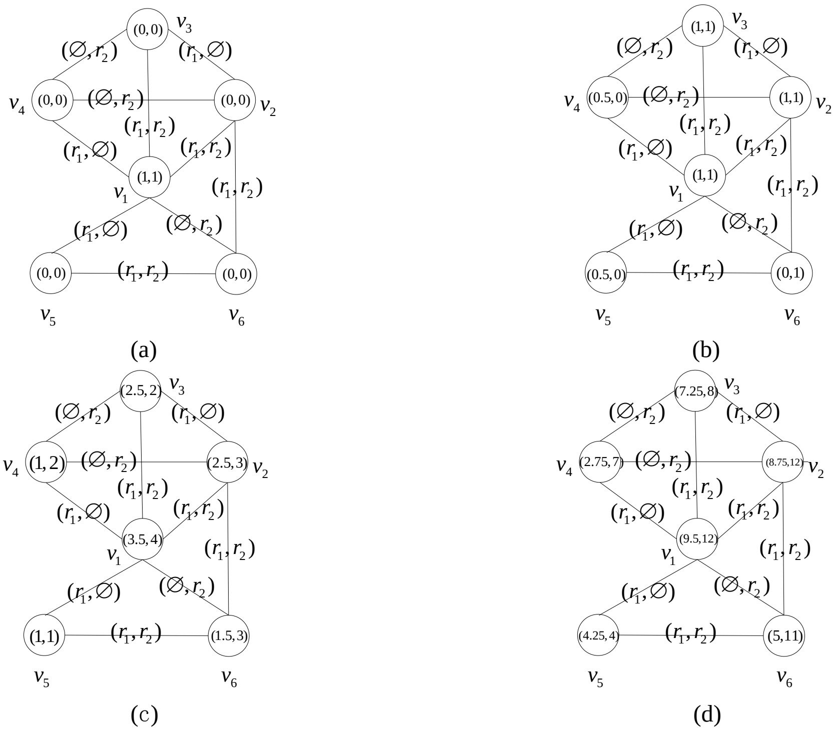

Figure 3. Signaling Spreading Process of Interactional Multi-Relationships Network with 6 Nodes

4. Node Importance Matrix Iterative Algorithm

The signaling spreading process in multi-relationship network could be described by matrix iterative operation. For node v_h in network $G(V, E)$, $\hat{k}_{v_h}^{r_i}$ represents degree of node v_h on relationship r_f, its definition is expressed as eq. (3).

$$\hat{k}_{v_h}^{r_i} = \sum_{v_l \in V} \hat{\boldsymbol{a}}_{hl}(r_i) \tag{3}$$

The node degree on relationship r_f indicates the number of nodes which signaling on relationship r_f can be directly reached.

Importance vector ***R*** can be conducted by adjacency matrix, importance matrix ***D*** can be constructed by adjacency matrix and importance vector ***R***. Based on above definitions, the steps of matrix iterative algorithm of importance matrix ***D*** are as follows:

(1) Generating the initial adjacency matrix $\hat{A}$ of multi-relationship network;

(2) Generating the initial importance vector ***R*** of every node by adjacency matrix $\hat{A}$;

(3) Constructing importance matrix ***D*** by importance vector ***R*** and adjacency matrix $\hat{A}$;

(4) Using matrix multiplication to simulate importance spreading, matrix multiplication mathematical equation as follows.

$$R_{i+1} = R_i \bullet D_i$$

(5) Comparing R_{i+1} with R_i to judge that the difference is whether or not more than set threshold value, which means that importance vector $\boldsymbol{R}$ 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

A dataset consisting of 30453 users in Douban.com [17] is selected to validate effect of the node importance matrix iterative algorithm proposed in this paper. The dataset network is composed of 30453 nodes and 55865 edges, there are friendship relationship and book comments similar relationship in the network, and the two kinds of relationship interrelates each other. So the dataset network is a typical multi-relationship online social network.

The dataset network is analyzed through the proposed algorithm and classical PageRank algorithm, the top 20 nodes in order of importance are shown in Table 1.

Table 1. Comparison of the Analysis Results of the Proposed Algorithm and Pagerank Algorithm

Rank	Our proposed algorithm	PageRank algorithm
1	No. 701	No. 6395
2	No. 7018	No. 3303
3	No. 1239	No. 701
4	No. 3356	No. 8220
5	No. 174	No. 7018
6	No. 209	No. 1239
7	No. 3549	No. 1299
8	No. 7132	No. 286
9	No. 3561	No. 3356
10	No. 4323	No. 7911
11	No. 3303	No. 174
12	No. 702	No. 13237
13	No. 2828	No. 286
14	No. 6461	No. 702
15	No. 4589	No. 209
16	No. 6939	No. 7132
17	No. 2914	No. 3561
18	No. 8220	No. 2828
19	No. 13237	No. 2914
20	No. 7911	No. 6939

From Table 1. we can see that though the sequence of the top 20 nodes is different for the two algorithms, there are only 4 nodes in the top 20 nodes sequence of our proposed algorithm not in the sequence of PageRank algorithm. If the top 100 and top 500 nodes of the two algorithms sequences are selected, the two sequences are completely covered. So for node importance analysis, the two algorithms can be considered to be concordant. But iteration times of our proposed algorithm is much less than times of PageRank algorithm. Consequently, node importance evaluation approach based on signaling spreading more applies to nodes importance analysis of multi-relationship online social network than traditional algorithms.

6. Conclusion

Based on signaling spreading, we proposed to use importance matrix iterative method to analyze node importance in multi-relationship online social networks. Through experiment analysis, it proved that comparing with existing node importance evaluation approaches, our proposed algorithm in this paper can better reflect node importance in multi-relationship online social networks, and iteration times of our proposed algorithm is much less. The research would contribute to improve efficiency and accuracy of public opinion analysis, community structure detecting, signaling spreading in multi-relationship online social networks.

References

[1] D. Watts and S. Strogatz, "Collective Dynamics of Small-World Networks", Nature, vol. 393, **(1998)**, pp.6684.
[2] A. Barabasi and E. Bonabeau, 'Scale-Free Networks", Scientific American, vol. 288, no. 5, **(2003)**.
[3] F. Jinqing, "Birth and Development Prospects of Network Science", Journal of Guangxi Normal University, vol. 3, no.5, **(2007)**.
[4] W. Jinshan and D. Zengru, "Complex Networks in Statistical Physics", Progress in Physics, vol. 24, no. 1, **(2004)**.
[5] Z. Tao, B. Wenjie and W. Binghong, "A Brief Review of Complex Networks, Physics", vol. 34, no. 1, **(2005)**.
[6] F. Jinqing, "Theoretical Research Progress in Complexity of Complex Dynamical Networks", Progress in Natural Science, vol. 17, no. 7, **(2007)**.
[7] H. Haibo and W. Xiaofan, "Analysis of Online Social Networks Based on Complex Network Theory", Complex Systems and Complexity Science, vol. 2, no.5, **(2008)**.
[8] Y. Gaohui, Y. Jianmei and Z. Mingang, "Research of QQ Group Friend Relationship Based on Complex Network", Journal of South China University of Technology, vol. 13, no. 4, **(2011)**.
[9] D. Callaway, J. Newman and S. Strogatz, "Network Robustness and Fragility: Percolation on Random Graphs", Physical Review Letters, vol. 85, no. 25, **(2000)**
[10] Y. Xia and J. Fan, "Efficient Attack Strategy to Communication Networks with Partial Degree Information", Proceedings of 2011 IEEE International Symposium, **(2011)**.
[11] L. Freeman, "A Set of Measures of Centrality Based on BeTweenness", Sociometry, vol. 40, no.1, **(1977)**.
[12] S. Milgram, "The Small World Problem", Psychology Today, 2, 1 **(1967)**.
[13] T. Opsahl, F. Agneessens and J. Skvoretz, "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths", Social Networks, vol. 32, no. 1, **(2011)**.
[14] H. Kim, J. Tang and R. Aderson, "Centrality Prediction in Dynamic Human Contact Networks", Computer Networks, vol.56, no.3, **(2012)**.
[15] Okamoto K, Chen W and Li X, "Ranking of Closeness Centrality for Large-Scale Social Networks", Lecture Notes in Computer Science, vol.50, no. 59, **(2008)**.
[16] H. Yanqing, L. Menghui and Z. Peng, "Community Detection by Signaling on Complex Networks", Physics Rev. E, vol. 78, no. 5, **(2008)**.
[17] Z. Qingqing, "Relationship between Scores and Tags for Chinese Books—In the Case of Douban Book", Chinese Journal of Library and Information Science, vol. 18, no. 4, **(2013)**.

Authors

Weihua Zhang, is currently an Associate Professor in the School of Computer Science and Engineering at Qingdao University. His main research interests include complex networks, web information retrieval and data mining.

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 PhD. 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.