

A Node Importance Analytical Approach for Multiple Relationships Online Social Network

^{1,2}Dong Suhui, ^{1,2}Sun Zhongting and ¹Xu Yonggang

¹China University of Mining and Technology, Xuzhou, China

²JiangSu United Vocational College, Xuzhou, China
xz_dsh@163.com

Abstract

Multiple relationships online social network node importance analysis was proposed based on signaling process, and considering the characteristics of multiple relationships which would interrelate with each other. The novel approach has been approved to be efficient by experiments described in this paper, and can improve efficiency of the analysis of public opinions, community structure detection and information propagation.

Keywords: *complex network, signaling process, node importance, multiple relationships online social network, matrix iteration*

1. Introduction

Online social networks, in which large groups of internet users with common interests or activities communicate and share resources with one another, make up a considerable portion of internet use [1]. Online social networks usually manifest as social networking services such as MYSpace or Facebook, content-sharing web sites such as YouTube or Flickr, online discussion groups such as Google Groups and all variety of message boards. Many researchers have hailed online social networks as egalitarian spaces that facilitate powerful information exchange, provide emotional support or mobilize users into collective action [2-4]. Understanding the complex social behaviors and group dynamics that take place in online social networks can provide considerable insight for social science.

Online social networks are typical example of complex network, using complex network theory to analyze online social network has become an effective way to understand accurately topology and evolution of online social network. The popularity of online social networks has also attracted many complex network researchers to study their topological characteristics [5-6], evolution models [7-9] and user behaviors [10-11].

We know very little about who drives or leads these social networks, or who serves as the most influential members of groups. A deep understanding of the communication traits and online social networks of leaders not only informs previous theories of leadership in organizational settings, or provides a new lens for identifying social interaction or information diffusion, but it provides insights what makes groups different than those in the realistic world.

Research on node importance in complex network originated from the field of social network analysis. In complex networks, those leaders of online social networks can be reflected through node importance. Finding the most important node in network was first proposed by Corley and Sha in 1982 [12]. In the early research, Freeman has done a lot of study in the sociology network, and pointed out the importance of node is equivalent to the significance with which the node connects with other nodes. Those typical topological characteristics include: degree method, closeness method and betweenness method, and so on.

Degree sorts the nodes by node degree, which is considered as a measure standard of node importance. The more edges connected with the node, the more important the node is. But only from the node degree the node importance can't be accurately judged in the network. If one node is with low degree, but it connects with some important nodes, the node may still be an important node in the network.

Closeness [13] reflects middle level of node in the network. The greater node's closeness is, the more important the node is. And it indicates that the node is much closer to the center of network. Closeness relies on the network topology. It can accurately discover the important nodes in centralized star networks, whereas can't apply to networks such as regular graphs and ER random diagram.

Freeman proposed Betweenness Centrality [14] in 1977. Betweenness Centrality means that the middle nodes have "a greater interpersonal impact" on the members at ends of paths. It is that, if a member is located in multiple shortest paths of other members, then the member is a core member with a larger Betweenness Centrality. But the topological characteristics has some limitations. It couldn't have an effective measure to evaluate the importance of edge nodes.

2. Node Importance Analytical Approach Based on Signaling Spreading

In online social networks, node importance can be defined as the impact of some node for information spreading ability in networks. For above-mentioned traditional node importance evaluation topological characteristics, degree only can represent information spreading ability of nodes on the local scale. Betweenness centrality and closeness only take nodes importance on topology location into consider, it would not reflect accurately the impact of nodes for information spreading. Therefore, it is highly necessary to research nodes importance analytical approach which is suitable for its characteristics of online social networks.

There are many social relationships, such as friendship, commerce, or others in an online social network. With regards to leaders, they can have different interests for different subjects. The communities of leaders tend to overlap. So it is inaccurate to detect and identify leaders in multi-relationship online social networks according to a kind of relationship alone.

With regard to describing multiple kinds of relationships and interrelations of complex system, layered network, was presented by Kurant *et al.* [15-16], a related but distinct model, where nodes and links in one layer represent one kind of relationship and interrelations and links between two different layers other kinds of interrelations.

Differing from previous studies, the aim of our study is to use signaling spreading, a new node importance analytical approach based layered network is proposed.

2.1. Signaling Spreading Mechanism

In complex networks, nodes can be converted into vectors of the algebraic space by using signaling spreading [17].

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 Figure 1.

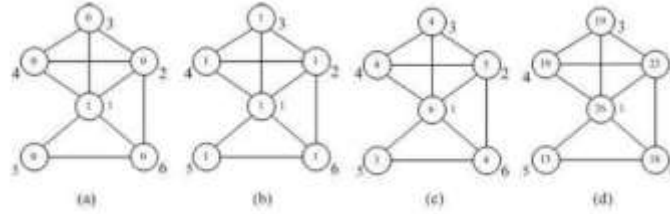


Figure 1. Signaling Spreading Process of Network with 6 Nodes

For the whole network with n nodes, n vectors will be got. The n vectors are projected to the algebraic space through standardization, and 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).

$$V = (I + A)^T \quad (1)$$

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 $V_i = (v_{i1}, v_{i2}, \dots, v_{in})$, then the V_i can be standardized as $U_i = (u_{i1}, u_{i2}, \dots, u_{in})$, here $u_{ij} = v_{ij} / \sum_j v_{ij}$.

2.2. Signaling Spreading in Multi-Relationship Network through Layered 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 through layered network. The i th vector component $s^{<v_i, v_j>}(r_f)$ 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 $(r_1, \dots, r_i, \dots, r_m)$ 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 Figure 2.

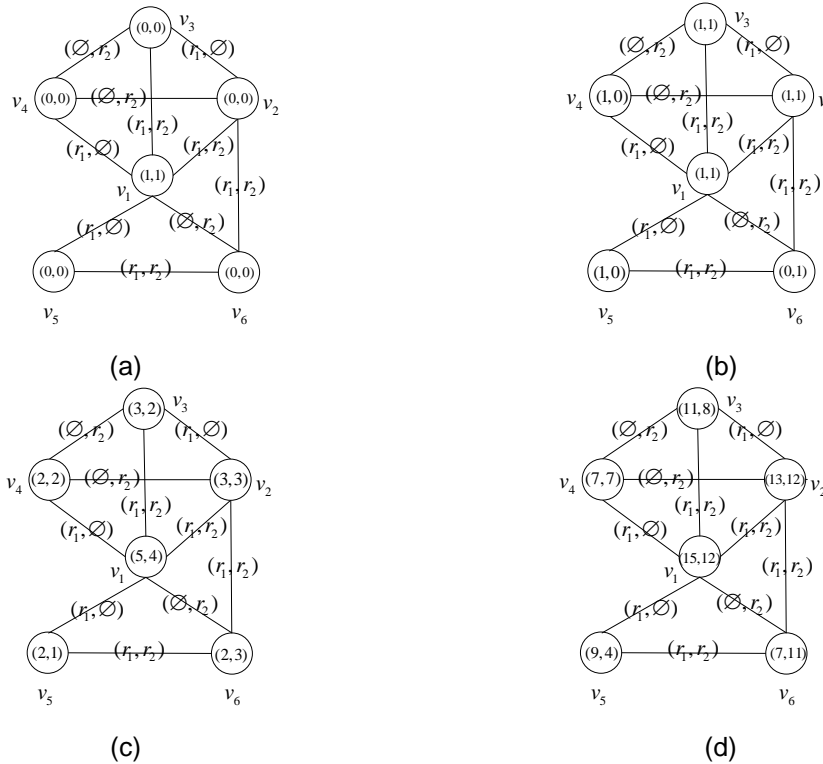


Figure 2. Signaling Spreading Process of Multi-Relationships Network with 6 Nodes

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

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

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

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

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:

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

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

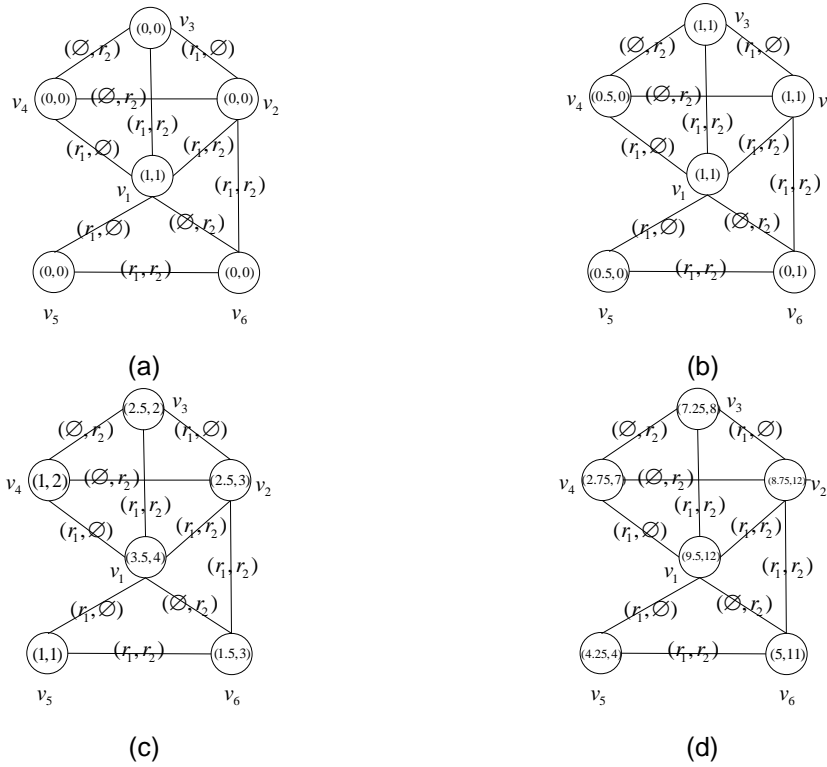


Figure 3. Signaling Spreading Process of Interactional Multi-Relationships Network with 6 Nodes

3. 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).

$$\hat{k}_{v_h}^{r_i} = \sum_{v_i \in V} \hat{a}_{hl}(r_i) \quad (3)$$

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

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

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

- (2) Generating the initial importance vector \mathbf{R} of every node by adjacency matrix \hat{A} ;
- (3) Constructing importance matrix \mathbf{D} by importance vector \mathbf{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 \mathbf{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.

4. Experiment and Analysis

A dataset consisting of 30453 users in Douban.com [18] 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 Page Rank 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.

If the difference of every node importance in two iterations is less than a certain threshold, at this time, the important degree of nodes is regarded as stable. The iteration times with different threshold to trend stability are shown in Figure 4.

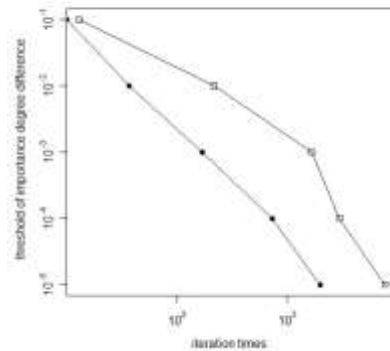


Figure 4. Iterative Times Comparison of Two Algorithms

From Figure 4 we can see that the convergence of our proposed algorithm is much better than the PageRank algorithm. In the case of the same importance difference threshold, iteration time of our proposed algorithm is much less than iteration times of PageRank algorithm. Consequently, node importance evaluation approach based on signaling spreading can more apply to nodes importance analysis of multi-relationship online social network than traditional algorithms.

5. 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.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 51204186).

References

- [1] T. Holtgraves and B. Lasky, "Linguistic power and persuasion", *Journal of Language and Social Psychology*, vol. 18, no. 2, (1999).
- [2] N. K. Baym, "Tune in, log on: Soaps, fandom, and online community", Sage Publications, New York, (2000).
- [3] H. Rheingold, "Smart mobs: The next social revolution", M.I.T. Press, New York, (2003).
- [4] B. Wellman and M. Gulia, "Smart mobs: Net surfers don't ride alone: Virtual communities as communities", M.I.T. Press, New York, (1999).
- [5] Y. Ahn, S. Han and H. Kwak, "Analysis of topological characteristics of huge online social networking services", *Proceedings of the 16th international conference on World Wide Web. ACM*, (2007).
- [6] S. Wu, J. M. Hofman and W. A. Mason, "Who says what to whom on twitter", *Proceedings of the 20th international conference on World Wide Web. ACM*, (2011).
- [7] F. Fu, X. Chen and L. Liu, "Social dilemmas in an online social network: the structure and evolution of cooperation", *Physics Letters A*, vol. 1, no. 371, (2007).
- [8] J. Leskovec and L. Backstrom, "Microscopic evolution of social networks", *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM*, (2008).

- [9] R. Kumar, J. Novak and A. Tomkins, "Structure and evolution of online social networks", Link mining: models, algorithms, and applications. Springer New York, (2010).
- [10] J. Jiang, C. Wilson and X. Wang, "Understanding latent interactions in online social networks", ACM Transactions on the Web, vol. 4, no. 7, (2013).
- [11] F. Benevenuto, T. Rodrigues and M. Cha, "Characterizing user behavior in online social networks", Proceedings of the 9th ACM SIGCOMM conference on Internet measurement. ACM, (2009).
- [12] H.W. Corley and D.Y. Sha, "Most vital links and nodes in weighted networks", Oper. Res. Letters, vol. 1, no. 21, (1982).
- [13] L. C. Freeman, "Centrality in social networks: ii. Experimental results", Social networks, (1979).
- [14] L. C. Freeman, "A set of measures of centrality based upon betweenness", Sociometry, vol. 40, (1977).
- [15] M. Kurant and P. Thiran, "Extraction and analysis of traffic and topologies of transportation networks", Physical Review E, vol. 74, no. 3, (2006).
- [16] M. Kurant, "Layered complex networks", Physical Review Letters, vol. 96, no. 13, (2006).
- [17] H. Yanqing, L. Menghui and Z. Peng, "Community detection by signaling on complex networks", Phys. Rev. E, vol. 78, no. 5, (2008).
- [18] 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



Dong Suhui, Dong Suhui is currently an Associate Professor in JiangSu United Vocational College, China. His main research interests include big data, network security, complex networks, cloud computing and data mining.



Sun Zhongting, is currently a lecturer in JiangSu United Vocational College, China. His main research interests include big data, Internet of things, intelligent information data processing, complex networks.



Xu Yonggang, is currently an Associate Professor in China University of Mining and Technology. His main research interests include signal processing, compressed sensing image processing, complex networks.