

A Method of discovering Key Nodes for Online Social Network based on Coritivity Theory

Jun Wang^{1, a}, Suozhu Wang^{2, b}

^{1,2} School of Management, Capital Normal University, Beijing, 100089, China

^a wangjuncnu@163.com, ^b wsz_wsz@163.com

Keywords: online social networks, core and coritivity, key node

Abstract. It has become well known that knowledge about key network members is essential for monitoring public opinion and controlling rumor propagation effectively in online social networks. According to global topology of online social network and coritivity theory, a coritivity-based method for finding key nodes in online social networks is proposed in this paper. In this method, the core of online social network is firstly found out, and then the importance of each node in core set is computed by using the traditional node importance measurement method so that further distinguish the importance of each node. Experimental results show that the method can discovery the key nodes in networks accurately, which is effective and feasible.

Introduction

With the rapid development of the Internet and intelligent mobile terminals, online social network (OSN) has become an important way for people to access and share information, make friends and entertainment, and it is at frontier of computer science and management science research. OSN has powerful information dissemination, spreading, access and sharing features and have attracted a large number of users. It have become available the most popular Internet application service at present. The massive nodes and complex relationship of each other are of great importance for research about identifying and evaluating significant nodes in social networks, such as if find out key nodes, it will contribute to monitor the social public opinion and control rumor propagation effectively during the dissemination of information in online social networks.

Since the advent of online social networks, there has been a lot of interest in finding important node and evaluating importance. Nowadays, researches about the method of node importance evaluation in online social networks mainly focus on social network structure and information diffusion cascade characteristics. The former algorithm basically includes Degree^{[1][2]}, Betweenness^[3], Closeness^[4], Page Rank^[5], HITS(Hyperlink-Indued Topic Searcher)^[6] and its related variant algorithms, which can be applied to different complex network calculation and evaluation. The latter one measuring the importance of nodes in the network use information diffusion cascade quantity, scale depth and other characteristics^[7]. The degree centrality of the node can measure the number of connection nodes of a node, which indicate whether the node is at the center of most nodes. The betweenness reflects the node in the network whether it is bridge in different subgroups, which can describe ability of the nodes to connect different subgroups. The closeness can represent the average distance between a node and other nodes. In social network specific topologies, these metrics contain nodes information about importance, prestige, status, etc. All above, these algorithms is calculating the importance of each node at first, and then find the node of top 10 through ranking the magnitude of the importance.

In the light of system theory, every system, existed in the real world, contains one or several entities that are essential for the whole system. If removed or destroyed, the system will be affected in all aspects, such as structure, stability. These entities are called cores in the system. In fact, the online social network is a large and complex system, so this paper constructs concepts about core and coritivity and introduces core nodes algorithm of online social network, providing a new idea and method for figuring out significant nodes.

Online social network and coritivity theory

Online social network is a social structure consisted of the set of social individuals and the connection between individuals in information network with the characteristics of large number of individuals and complex network structure. Through graph theory and social graphs to visualize the social network.

Social network graph

In online social networks, users and connections between them can be visualized as nodes and lines separately. Therefore, graph $G = \langle U, L \rangle$ can represent social network abstractly, where U denotes the set of users in social network, and node $u \in U$ represents an individual, L denotes the set of connections between users, including directed edge sets or undirected edge sets, and edge $e \in L$ represents interpersonal relationships in social networks, such as friendship and partnership.

Coritivity and core in social network

Constructing concepts about coritivity and core of social networks requires some knowledge about graph theory, related definitions listed as below:

Definition 1 (Node connection). In undirected graph $G = \langle U, L \rangle$, $u, v \in U$, if there is a path between u and v , then u and v are said to be connected.

Definition 2 (Connected graph). In undirected graph G , if $\forall u, v \in U$, and u and v are connected, then G is called a connected graph.

In the definition 1, connection relationship between nodes is the equivalence relationship about U . According to the equivalence relationship, U can be classified, if and only if u and v are connected and in the same class.

Definition 3 (connected components). Let U_1, U_2, \dots, U_p be the equivalence class of U , sub-graph $G[U_1], G[U_2], \dots, G[U_p]$ is called the connected components of G , and the number of connected branches is denoted as $\omega(G)$.

Definition 4 (Vertex cut set). Let G be an undirected graph, $U' \subseteq U$, if $w(G - U') > 2$, then U' is called vertex cut set of G .

Based on the coritivity theory and core ideology in system theory, there are also cores in the social networks, and they are key nodes. Core nodes play a key role in maintaining stability or information dissemination in the social network. To measure core nodes' importance quantitatively, define coritivity and core in social network as below:

Definition 5 (Coritivity and core of online social network). Let $C(G)$ represents a set of all vertex cut sets in online social network graph G . Assuming that the total number of social network users is $|U(G)| = n$, then the coritivity of graph G , $h(G)$, is defined as

$$h(G) = \max \{ \omega(G-S) - |S|; S \in C(G) \} \quad (1)$$

where $\omega(G-S)$ denotes the number of components of graph $G-S$, $|S|$ denotes cardinality of vertex cut set. Furthermore, if $S^* \subseteq C(G)$ and satisfies

$$h(G) = \omega(G-S^*) - |S^*| \quad (2)$$

S^* be called a core of graph G .

In the social network graph G , the coritivity is unique, while the core is not in that could be one or more cores in a social network graph. The coritivity reflects the macrocosm connectivity of social network graphs. The bigger the coritivity is, the worse the network connectivity, and the smaller the coritivity, the better the network connectivity. After finding out cores in online social network, calculate the importance of them by the traditional importance measure method, and key nodes can be sorted further.

Calculation Method of coritivity and core in social network

Equivalent form of coritivity calculation formula

In [8] authors study and give a method for computing coritivity of arbitrary system. In this paper, the method of [8] is used to compute coritivity and core of online social network.

Definition 6. Let G be an undirected graph of order P , with $\omega(G)$ connected components, then $r(G)$ is said to be the rank of graph G , where $r(G) = P - \omega(G)$.

Definition 7. Let G be a simple undirected graph, with p nodes and q edges, then $A(G) = [a_{ij}]_{p \times q}$ is called the complete incidence matrix of G . If edge e_j is associated with the vertex v_i , then $a_{ij} = 1$. If not, $a_{ij} = 0$.

Theorem 1. Let G be a connected graph of order P , if and only if the rank of complete incidence matrix of graph G is $P-1$.

Theorem 2. The rank of graph G is equal to rank of complete incidence matrix of G .

The proofs of Theorem 1 and theorem 2 may be referred to the relevant literature.

According to theorem 1 and theorem 2, the coritivity $h(G)$ can be work out by calculating the rank of graph $G-S$, which is also the rank of the complete incidence matrix of graph $G-S$, where S denotes vertex cut set. As a result,

$$h(G) = P - \min\{r(G-S) + 2|S|; S \in \mathcal{C}(G)\} \quad (3)$$

Eq. (3) simplifies the complexity of calculus, through transforming the computation of the coritivity of a social network graph into the rank of the complete incidence matrix.

Algorithm of coritivity and core

Based on the analysis above, calculation specific steps of coritivity and core are as follows:

Step 1: In social network graph $G = \langle U, L \rangle$, $|U(G)| = P$, mark nodes as $U = \{U_1, U_2, \dots, U_i, \dots, U_p\}$, U corresponding to a natural number set $N = \{1, 2, \dots, I, \dots, P\}$.

Step 2: Construct complete incidence matrix $A(G)$ of graph G .

Step 3: Work out all nontrivial subsets N_{ij} of N , where i denotes there are i elements in N_{ij} , j denotes the j -th set with i elements.

$$N_{ij} = \{n_{ij1}, n_{ij2}, \dots, n_{ijk}, \dots, n_{iji}\}, n_{ijk} \in N \quad 1 \leq k \leq i, 1 \leq i \leq P-1, 1 \leq j \leq C_p^i$$

Step 4: In the complete incidence matrix $A(G)$, for each subset N_{ij} , cross out the row and column corresponding to the n_{ijk} , in that cutting out a node should cross out all the edges connected to the node in a network graph, in order to form complete incidence matrix $A(N_{ij})$, and then calculate the rank $r(N_{ij})$ of $A(N_{ij})$ by mod 2 algebra.

Step 5: If $0 < r(N_{ij}) < P-i-1$, then $b_{ij} = r(N_{ij}) + 2i$.

Step 6: Calculate $B = \min(b_{ij})$.

Step 7: $h(G) = P-B$.

As a result, N_{ij} is the core on the condition that satisfied the Eq. (3). The algorithm can not only figure out the coritivity, but also the core of the graph G .

Experiment and evaluations

Zachary network, a classical data set in social network analysis field, describes the relationship

between the 34 members of an American university Karate Club, which contains 34 nodes and 78 edges (Figure 1), and the data set has been used for analysis of a lot of algorithms.

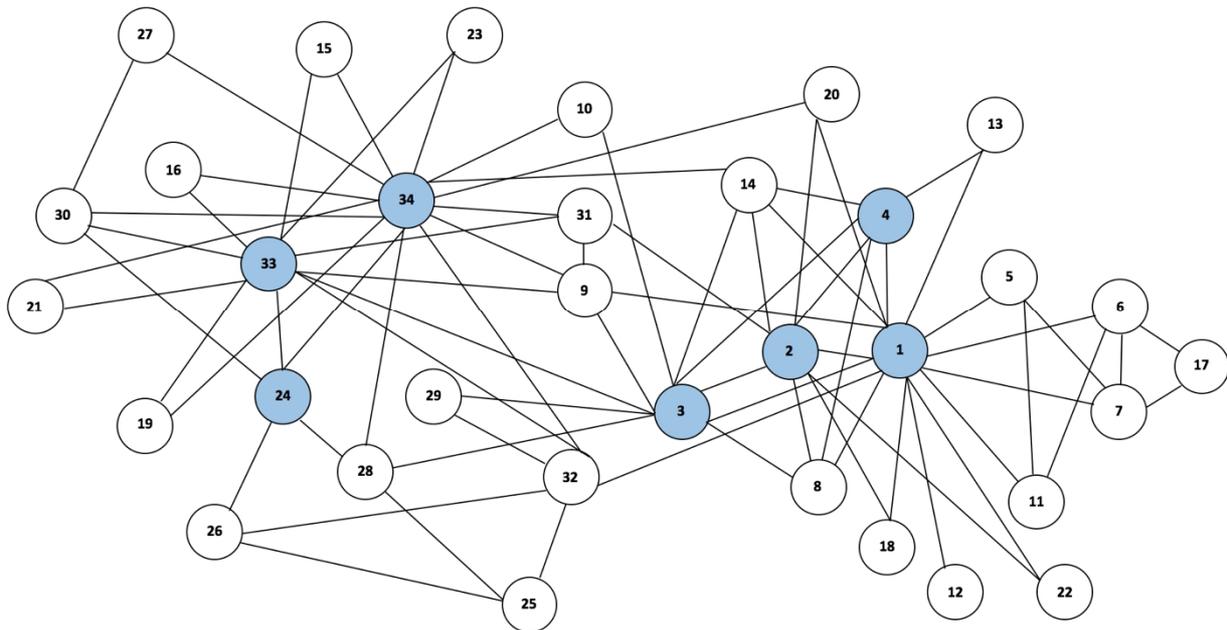


Figure 1 Karate Club Social Network Graph

In this paper, the algorithm is implemented on Windows 7 platform by using Python 2.7, and the algorithm is validated by the data set. The experimental result shows that the coritivity of the karate club membership network is 10, with unique core set {1, 2, 3, 4, 24, 33, 34}. To Compare the importance of each node in the core set, degree method is used for calculating nodes importance, and compared with Page Rank algorithm, as shown in table 1.

Table 1 Rank of nodes

Nodes	Degree	Rank by Degree	Page Ranks	Rank by Page Rank
1	16	1	0.100878646	1
2	9	5	0.052872059	5
3	10	4	0.057075052	4
4	6	6	0.035861518	7
24	5	7	0.031520683	8
33	11	3	0.071663214	3
34	14	2	0.097001115	2

According to Table 1, Node 1 and Node 34 is the core nodes in the Karate Club membership network, which coincide with the two nodes are community leaders in actual situation, and the top 7 important nodes coincide with the experimental result of Page Rank as well. Experimental result demonstrates that the method proposed in this paper can calculate the key node or key node set in the social network accurately, and it is effective and feasible.

Conclusion and future work

It is online social network that plays a growing more important role in information dissemination, therefore how to identify their key nodes is extremely significant for controlling information diffusion effectively. Focusing on the global characteristics of online social networks, this paper put forward a method of figuring out core nodes based on the coritivity theory in system theory. Experimental result demonstrates that the method is feasible. However, the computation is large and time-consuming with the number of nodes and the relationship between nodes is

numerous in the online social network. It requires that adopt big data algorithm and design distributed algorithm to improve the method efficiency for figuring out core nodes in networks and applied to large-scale social network analysis in the future research.

Acknowledgements

This research was supported by National Natural Science Foundation of China (grant No. 61371177) and by the Science and Technology Plan of the Beijing Education Committee (grant No. KM201510028019).

Reference

- [1] Callaway D, Newman J, Strogatz S, Watts D. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, Vol. 85-25(2000), p. 5468-5471.
- [2] Reza Zafarani, Mohammad Ali Abbasi, Huan Liu: *Social Media Mining An Introduction*, The People's Posts and Telecommunications Press, Beijing(2015).
- [3] Hao-xiang Wang, Shan Zeng, Hui-yang Liu. An Importance Analytical Approach for Online Social Network. *Journal of Shanghai Jiaotong University*, Vol. 47-7(2013), p. 1055-1059.
- [4] Okamoto K, Chen W, Li X. Ranking of closeness centrality for large-scale social networks. *Lecture Notes in Computer Science*, Vol. 5059(2008), p. 186-195.
- [5] Page L, Brin S, Motwani R, Winograd T. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report(1999).
- [6] Kleinberg J M, Authoritative sources in a hyperlinked environment. *Journal of the ACM*, Vol. 46-5(1999), p. 604-632.
- [7] Dong Li, Zhi-ming Xu, Sheng Li, Ting Liu, Xiu-wen Wang. A Survey on Information Diffusion in Online Social Network. *Chinese Journal of Computers*, Vol. 37-1(2014), p. 189-206.
- [8] Suo zonglin, Yang yinsheng, Gaoyixin. A General algorithm for the system coritivity. *Journal of System Engineering and Electronics*, Vol. 3(1995), p. 27-28.