PAPER • OPEN ACCESS

Research on the discovery of opinion leaders in social networks

To cite this article: Chufeng Zhou *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **563** 032009

View the article online for updates and enhancements.

Research on the discovery of opinion leaders in social networks

Chufeng Zhou[1,a] , Xinxin Guan[1,b], Yeli Li[1,c], and Qingtao Zeng[1,d]

[1]School of Information Engineering, Beijing Institute of Graphic Communication,

Beijing 102600, China

[a]576213376@qq.com, [b] 1145512971@qq.com, [c] liyl@bigc.edu.cn,
[d]zengqingtao@bigc.edu.cn

Abstract. Since entering the era of Wed 2.0, social networks have been an indispensable part of human life. At present, these social networks are the main source of information dissemination and alienation, and they play a key role in the process of dissemination and are the central users of social networks. Central users have an incredible ability to communicate in their community, and their knowledge and experience of things have a great impact on other users in the community. This research mainly uses the analysis method of social network to analyze the community and find the central users in the community. We use Louvain algorithm to identify the community structure in social networks, and on this basis we use PageRank algorithm to analyze the central users in each community.

1. Introduction

In the era of Wed2.0, a large amount of data will be generated every day. User data generated in social networks are especially large, such as Facebook, Twitter, Tumblr, Instagram, etc. These social networks provide an opportunity to interact with known things and human beings in another unknown world. They are the main source of information dissemination and alienation. At the same time, social networks also provide a level for companies. In the process of disseminating commodities and new products, one kind of user has a great influence on the adoption process and decision-making of other users. Lazarsfeld et al. introduced this phenomenon at seminars in 1940 and 1950 (Gold, Katz, Lazarsfeld, & Roper, 1956), and called these users opinion leaders. Lokesh Jain and Rahul Katarya (2018) used an improved firefly algorithm to find local and global opinion leaders. Mak (2008) used a game theory-based approach to determine the invalid or weak results of opinion leader-follower association. Li and Du (2011) proposed a book-based approach. The BARR model of body, including blog content, authors, readers and their relationships, identifies hot topics.

This paper introduces the problem of finding community and community leaders in social networks. In order to solve these problems, firstly, we divide social networks into several communities by algorithm. On this basis, we use PageRank algorithm to identify opinion leaders in the community.

2. Related Work

In this section, we will discuss some concepts related to social networks to clarify our research. Social network is a social structure consisting of social individual aggregation and connection among individuals on information network. With the rapid development of Internet technology, the number of netizens has increased exponentially. User-generated content (UGC) has been growing continuously.

It is especially evident in social network. Facebook users share more than 4 billion things every day, and Twitter processes more data every day. Over 340 million Tumblebr bloggers post 27,000 new posts per minute, and the value of these data is far greater than the platform itself, so social network analysis is becoming more and more important.

Social network can be defined clearly according to the non-directed graph G = (V, E). V and E are the set of nodes and links in the network. There are several important concepts involved, namely, degree centrality, proximity centrality and intermediary centrality. We will introduce these concepts separately.

2.1 Degree Centrality

Degree Centrality refers to the sum of connections between a node and other nodes, which can indicate the degree of cohesion of a node in the network social circle. The larger the size of a node, the higher the centrality it occupies. But in reality, there is a directional connection, that is, directional connection, so the concepts of in-degree and out-degree come into being.

In-degree realizes a person's degree of concern, refers to the other nodes want to form an associated object with it, which reflects the attraction of the node.

Out-degree represents the degree to which a person pays attention to others, meaning that the node strives to associate with others, and people who has high out-degree can obtain rich information from other nodes in the network.

2.2 Closeness Centrality

The sum of distances from a point to another point is calculated. The sum of distances shows that the shorter the path from this point to all other points, that is, the nearest point to all other points. Bavelas (1950) defined proximity to centrality as the reciprocal of distance:

$$C(x) = \frac{1}{\sum_y d(y,x)} \tag{1}$$

The sum of the shortest distance between a point and other points can be normalized to get a number between (0,1). The larger the number, the higher the proximity of the point to the center. It can be imagined that when the molecule in the formula tends to infinite infinity, the value of C tends to zero, so when the distance between a point and all other points is very large, that is to say, the point is not in the central position, then its proximity to the center tends to zero.

Similarly, in a directed social network, two concepts, in-closeness centrality and out-closeness centrality, will be discrepant from the analysis results of proximity centrality.

(1) In-closeness centrality

In-closeness centrality measures the ease at which other points reach this point by calculating the edge toward a point. The higher the In-closeness centrality of a point, the easier it is to point to other points.

(2) Out-closeness centrality

Out-closeness centrality refers to the ease with which a point reaches other points, represented by the reciprocal of the sum of the shortest distances from one point to the other. The larger the out-closeness centrality, the easier it is to go to other points.

Therefore, integration is expressed in approaching centrality, and radiation is expressed in approaching centrality.

2.3 Between Centrality

We calculate the number of shortest paths through a point. The greater the number of shortest paths passing through a point, the higher its betweenness centrality. If a large social network contains several groups, then people with high centrality will play the role of connecting these groups.

3. Finding and Realizing Opinion Leaders in Community

This section will discuss the implementation of partitioning in social networks and the algorithm for finding opinion leaders in social networks. Firstly, we use Louvain algorithm to find out the communities that are built in the social network with modular gains. Next, we use the Pagerank algorithm to identify opinion leaders in the community.

3.1 Community Partition Discovery Algorithms

Louvain algorithm is a community discovery algorithm based on graph data. The algorithm scans all the nodes in the data, and measures the benefits of modularity brought by adding the node to the community where the neighbor nodes are located, aiming at each node traversing all the neighbor nodes of the node. Neighbor nodes corresponding to the maximum benefit are selected to join their communities. This procedural repetition guides the community ownership of each node to remain unchanged. Modularity calculation such as formula 2:

$$Q = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \sigma(c_i, c_j) \tag{2}$$

A_{ij} represents the weight of the edges between node i and node j. k_i is the sum of the weights of all the edges connected with node i. c_i is the cluster of node I. & (c_i, c_j) means that if node i and j are in the same cluster, then return 1, otherwise return 0.

Modularity gain calculation such as formula 3:

$$\Delta Q = [\frac{L_{in} + k_{i,in}}{2m} - (\frac{L_{tot} + k_i}{2m})^2] - [\frac{L_{in}}{2m} - (\frac{L_{tot}}{2m})^2 - (\frac{k_i}{2m})^2] \tag{3}$$

We fold the communities formed in the previous steps, and fold each community into a single point. We calculate the edge weights between these newly generated "community points" and the sum of the edge weights between all the points in the community, for the next round of calculation.

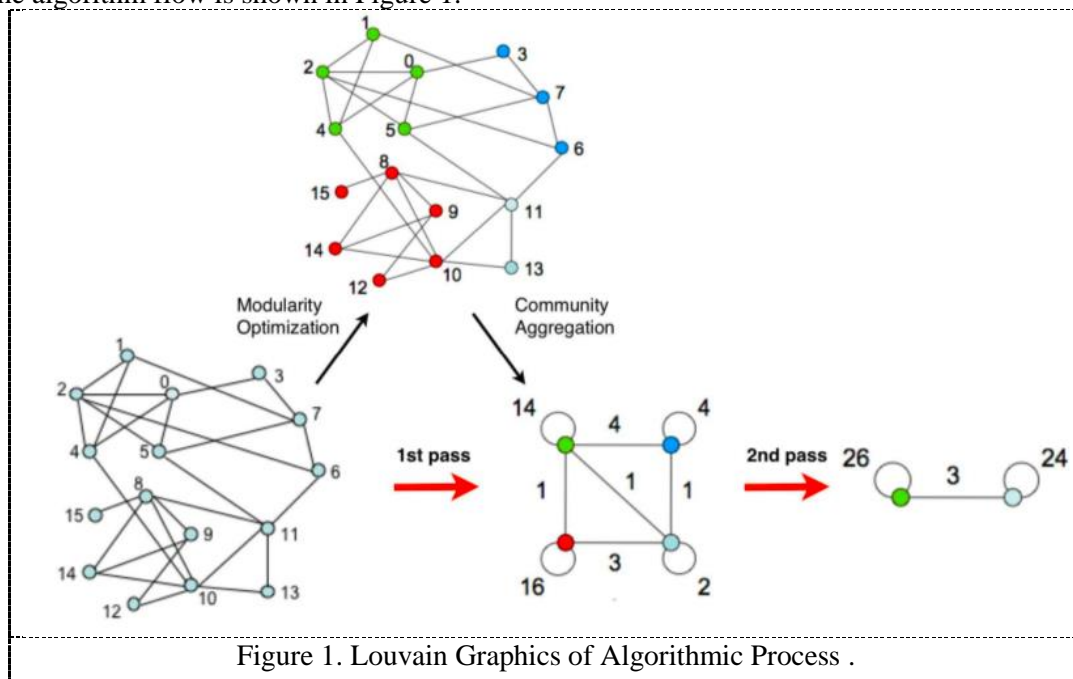The algorithm flow is shown in Figure 1.

Figure 1. Louvain Graphics of Algorithmic Process .

3.2 PageRank algorithm for finding community opinion leaders

PageRank algorithm was first proposed to solve the problem of page ranking. The main idea is: if a page is linked to a large number of other pages, it shows that the page is more important, that is, the PR value of the page is higher; if a page with a high PR value is linked to another page, the PR value

of the linked page will be relatively higher. Based on this feature, we can apply PageRank algorithm to find community opinion leaders.

For social networks, they have similar structures that connect with each other. Social networks can also be regarded as directed graphs. The links between pages point to each other, while the users of social networks point to each other by mutual attention. So in general, the PR value of a social network node can be calculated by the following formula 4:

$$PR\left(p_i\right) = \alpha \sum_{p_j \in Mp_i} \frac{PR\left(p_j\right)}{L\left(p_j\right)} + \frac{(1-\alpha)}{N} \tag{4}$$

Among, M_{Pi} is a set of nodes that pay attention to the node p_i, L (p_j) is the number of nodes p_j, N is the total number of nodes and α is generally taken as 0.85 after experiment. According to the formula 4, we can find opinion leaders in the divided communities, calculate the PR value of each node, after iteration, that is the final result.

4. Conclusion

In this study, the louvain algorithm was implemented to partition the social network, and the community opinion leaders were found through the PageRank algorithm in the partition. For the two algorithms, they both have the characteristics of fast and accurate, which can better complete the research objectives. Through this experiment, the selection of community opinion leaders has been successfully completed, which is of great help to the future research on social network analysis.

Acknowledgement

References

[1]   Zhao, Y., Kou, G., Peng, Y., & Chen, Y. Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective.Information Sciences 2018.

[2]   Zhu, M., Lin, X., Lu, T., & Wang, H. Identification of opinion leaders in social networks based on sentiment analysis : Evidence from an automotive forum. Advances in Computer Science Research,2016.

[3]   Aghdam, S. M., & Jafari Navimipour, N. Opinion leaders selection in the social networks based on trust relationships propagation. Karbala International Journal of Modern Science, 2(2), 88–97. 2016.

[4]   Aleahmad, A., Karisani, P., Rahgozar, M., & Oroumchian, F.. OLFinder: Finding opinion leaders in online social networks. Journal of Information Science. 2016.

[5]   Jiawei Zhang,Philip S. Yu:Broad Learning: An Emerging Area in Social Network Analysis.In ACM SIGKDD,2018.

[6]   J. Zhang, L. Cui, P. Yu, and Y. Lv. Bl-ecd: Broad learning based enterprise community detection via hierarchical structure fusion. In CIKM, 2017.

[7]   Y. Zhao, E. Levina, and J. Zhu. Community extraction for social networks. Proceedings of the National Academy of Sciences, 2011.

[8]   J. Zhang and P. Yu. Pct: Partial co-alignment of social networks. In WWW, 2016.

[9]   Lokesh Jain, Rahul Katarya:Discovery Algorithm of Opinion Leaders for Network Community Based on Pagerank 。Computer Engineering 2018

[10] Moradabadi B, Meybodi M R. Link Prediction inStochastic Social Networks: Learning Automata Approach[J]. Journal of Computational Science, 2017.

[11] LAPPAS T，TE R ZI E，GUNOPULOS D, et al. FindingEffectors in Social Network [c] //Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York，USA: ACM Press, 2010: 1059-1068

[12] Li, F., & Du, T. C. Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs. Decision Support Systems, 51(1), 190–197. 2011.

[13] Gold, D., Katz, E., Lazarsfeld, P. F., & Roper, E. Personal influence: The part played by people in the flow of mass communications. American Sociological Review. 1956.

[14] Mak, V. The emergence of opinion leaders in social networks. SSRN Electronic Journal.2008.

[15] Luo, L., Yang, Y., Chen, Z., & Wei, Y. (2018). Identifying opinion leaders with improved weighted LeaderRank in online learning communities. International Journal of Performability Engineering. doi:10.23940/ijpe.18.02.p1.193201.

[16] Ma, N., & Liu, Y. (2014). SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork. Expert Systems with Applications. doi:10.1016/j.eswa.2013.08.033

[17] Aleahmad, A., Karisani, P., Rahgozar, M., & Oroumchian, F. (2016). OLFinder: Finding opinion leaders in online social networks. Journal of Information Science. doi:10. 1177/0165551515605217.

[18] Heidemann, J., Klier, M., & Probst, F. (2010). Identifying key users in Online Social Networks: A PageRank based approach. In Proceedings of the 31st international conference on information systems. doi:10.1016/j.infsof.2008.09.005

[19] Li, F., & Du, T. C. (2011). Who is talking? An ontology-based opinion leader identifi-cation framework for word-of-mouth marketing in online social blogs. Decision Support Systems, 51(1), 190–197. doi:10.1016/j.dss.2010.12.007.

[20] Li, Y., Ma, S., Zhang, Y., Huang, R., & Kinshuk (2013). An improved mix framework for opinion leader identification in online learning communities. Knowledge-Based Systems, 43, 43–51. doi:10.1016/j.knosys.2013.01.005.