

The Data Analysis of Tencent Micro-blog User

Zhanshen Feng*¹, Shaohua Tao², Chun Yang³

^{1,2}Department of Information Engineering, Xu Chang University, Xu Chang 461000, Henan, China

³ Guangdong Power Grid Corporation Limited Information Center, Guangzhou, China 510080

Abstract: Micro-blog has been an important social network platform in recent years. With the advances in information influence and propagation ability, the date of micro-blog has attracted the attention of scholars of all kinds. More and more people are interested in studying microblogging from various aspects, such as user motivation, information propagation, text analysis, user network structure. This paper is about the data analysis of tencent microblogging user. Then get the data information about user's friends and audiences from the database. The simulation results show that fans distribution of user is power-law and average length is smaller, which indicate that user are of local network properties.

Keywords: Micro-blog; Data analysis; User; Fans

I. Introduction

Short for weibo, namely (Microblogging), is between different users can realize information sharing, information dissemination and access to useful information platform, users can through the WEB, WAP and various client form individual community, use about 140 words of text messages, and realize the real-time share[1,2,3,4,5].

Now the status quo of weibo is not only a social platform, we can draw news from weibo, share price fall, sense of the common life and so on. Because of weibo real-time of updating, high efficiency and convenience, people almost can receive the practical news around us at any time. Weibo broadcasting information, make information widely and rapidly. At present, weibo has been in the spread of infectious diseases, detection of news events, time orientation, company looking for potential users against all have the positive role, in addition, through the emerging issues of the microblog to predict the future trend of the development [2, 3]. Weibo users in the system through "attention" and "concern". Users can take the initiative to "focus on others", become the fans of others, can get the information of the other party without the consent of the other. Users via mobile phones, instant messaging, and web of personal weibo updating, if fans interested in the topic can immediately after receiving and forwarding to post evaluation[6,7,8,9,10].

Grabbing the tencent weibo users, this paper analyzes the user on the number of fans, friends, average shortest path, intimacy and other network parameters. Simulation indicates that the user on the number of fans present power-law distribution, average path small wait for a characteristic, this also shows that weibo has local characteristics. After the microblogging network development play a role, but also for the microblogging network application in the practical guiding function, positive impact to people's life.

II. Network Parameters

2.1 Degree and degree distribution

Directed graph, the degree of a node is divided into and out of the degrees. Into the degree is the number of edges into the node, the degree of starting from the node of the number of edges. Node degree distribution $P(k)$ said the number of nodes in the network medium for k proportion of network summary points. $P(k)$ are given an arbitrary choice of the probability of nodes linked by k edge. Degree distribution function to reflect the statistical characteristics of complex systems.

Type of degree distribution network is divided into five kinds: (1) type power law degree distribution; (2) the exponential distribution; (3) the type of the power-law degree distribution; (4) of truncated exponential distribution; (5) gaussian distribution. While weibo belongs to another - poisson (Pofsson) distribution, namely with characteristics of random network and small world network distribution type [10].

2.2 Clustering coefficient

Clustering coefficient $C(i)$ is described and the first node connection 1 probability of nodes are connected. Is a node between two friends is also a friend of the probability. Sense, at the edge of the connection node if the k_i of the i -th a node degrees, by the k_i subnet consisting of a neighbor node, the actual number

of edges $E(i)$ with all k_i nodes to connect fully when the total number of edges of $k_i * (k_i - 1) / 2$ ratio is defined as the clustering coefficient of node i , namely

$$C(i) = \frac{2E(i)}{k_i * (k_i - 1)}$$

$C(i)$ the closer to 1, indicate that the node point near the has a tendency to form a smooth dough. A large clustering coefficient, which indicates that the user in an interconnected network, can more quickly than low coefficient of users know events in the network.

2.3 Average path length

In the network, the path length l refers to the distance between two nodes, i.e. starting from a node i , after connected with its point, gradually to "walk" to another node the number of the edge of the distance traveled by j . Average shortest path length refers to the average of the shortest path between two points in a network. Node number is n undirected network, its average l is the path to the average shortest distance between network nodes

$$l = \frac{1}{n(n+1)/2} \sum_{i \geq j} d_{ij}$$

In the formula, d_{ij} is the shortest distance from node i to node j .

III. The Simulation Analysis

This article, we use tencent's API grabbed 100000 users, in the database, capturing 20334 weibo users of xuchang region, in front of each user has a different node id, the id. The connection relationship between the user via the user to determine the relationship between tables, and analyzed the number user audience to listen to and xuchang region.

3.1 Degree distribution

Weibo users in the system through "attention" and "concern". Users can take the initiative to "follow" other people, become the "fans" of others, to receive updated information, without the consent of the other. Similarly, other users can also be "concern", become your "fans", this process is a two-way reversible if the user use a network node, focus on between the user and the relationship nodes, weibo users and their relationship network is directed to network. Connect the more the number of edges, on behalf of the greater the degree of nodes, the greater role in the network. Degree (degree) In the user the number of fans, shown as Fig. 3.

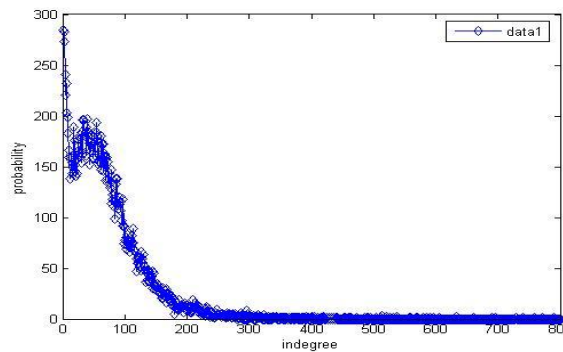


Figure 3 The degree distribution

Can be seen from the figure 3, in xuchang, henan tencent weibo user 80% users in 200 the number of audience, it is not hard to see such a user of in xuchang area, much of their main play the role of the news spread in the network. And only a tiny number of users to take a lot of fans, present power-law distribution network. The number of degrees on behalf of the user's friends, user's friends number after reaching a certain number, the present saturation state, shown as Fig. 4.

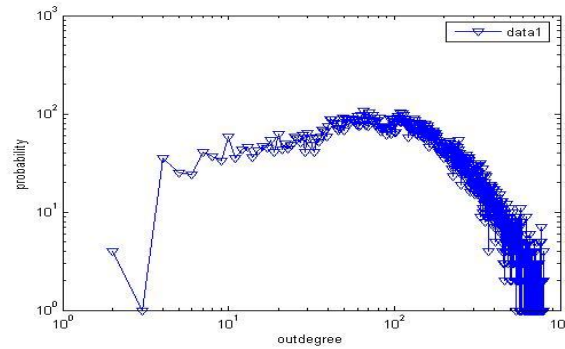


Figure 4 The user's degree distribution

3.2 Average path length

Without having the right to network, the path length refers to the two nodes from other nodes, after connected with its point, gradually to "walk" to another node by the number of the edge of the distance. The simulation of all the nodes path length is shown as Fig.6.

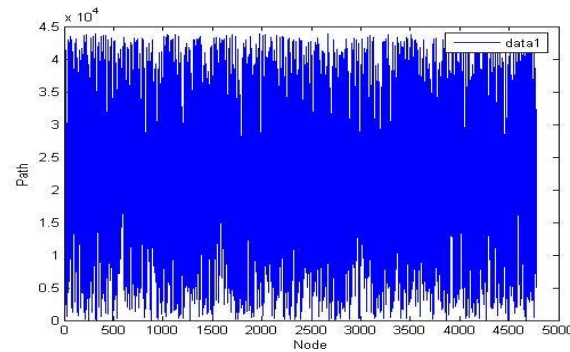


Figure 6 The path length

By calculating the average path length between the authors tencent weibo users $L = 1.76$. Because this project only henan xuchang users, they compare closely links in the network, so the relatively small average path length.

3.3 Cluster coefficient

In undirected network, cluster coefficient C_i describes the possibility of connection between neighbor nodes is connected to the same node in the network between the two nodes connected probability, which is a node between two friends is the probability of a friend. The closer to 1, said the node point near the has a tendency to form a smooth dough. Cluster coefficient is an important parameter to investigate the degree of collectivization of network. The following figure 7 for the microblogging network clustering coefficient probability distribution:

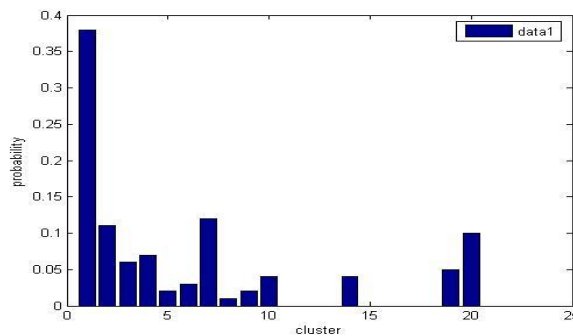


Figure 6 node aggregation coefficient of probability distribution

The figure 7 shows that in the microblogging network a user's the probabilities of two friends and friends are very small. The reasons for this situation is tht the database did restrictions on the user's information. In gathered coefficient equals 0.04, for example, the number of users has accounted for 38% of all users. In

gathered coefficient equals 0.5, the number of points in both sides of the user can see clearly from the figure of 80% of users gathered coefficient is less than 0.5.

IV. Conclusion

In this paper, first of all, the parameters of network are introduced, such as degree distribution, average path length, clustering coefficient and rate of attention, etc. Henan authors tencent weibo users against the grab, details the grab the API interface and the methods of the data. And tencent weibo user data of xuchang area to carry on the simulation analysis. Xuchang region tencent weibo users on the number of fans present power-law distribution, the simulation showed that 80% of users have fewer fans, while 20% of users have a large number of fans, such as a certified individuals, businesses, etc. Smaller average path of the whole network, shows that represented by the regional network of the authors has close between the user and contact more closely. Due to grab the user is limited, while the average path is small, but gathered coefficient is not too big, that is to say, has not yet been gathered to form a smooth dough. Therefore, the next step we will expand the data capture range, make a deeper analysis and research.

Acknowledgements

This work was supported by the Foundation Name under Grant No. (122102210544 and No.12A510022 and No. (132300410349)) and natural foundation fund (No. U1304403).

References

- [1]. Haewoon Kwak, Changyun Lee, Hosung Park, and Sue Moon. What is Twitter, a Social Network or a News Media? International World Wide Web Conference Committee(IW3C2), ACM,2010.
- [2]. M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the Flickr social network. In Proc. of the 18th international conference on World Wide Web. ACM, 2009.
- [3]. D. Zhao and M. B. Rosson. How and why people twitter: the role that micro-blogging plays in informal communication at work. In Proceedings of theACM 2009 international conference on Supporting group work. ACM, 2009.
- [4]. M. E. J. Newman and J. Park. Why social networks are different from other types of networks. Phys. Rev. E, 68(3):036122, Sep 2003.
- [5]. J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In Proc. of the third ACM international conference on Web search and data mining. ACM, 2010.
- [6]. Lars Backstrom, Jon Kleinberg, Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook. arXiv:1310.6753v1 [cs.SI] 24 Oct 2013.
- [7]. Xiaoke Xu, Xiaofan Liu, New power of network science development: Big Data and Crowdsourcing. Journal of Electronic Science and Technology University, 42(6):802-805,2013.
- [8]. Jinqing Fang, Big data challenges and opportunities under the impact of network science and engineering. Natural,35(5):345-354,2013.
- [9]. Z.Q.Jiang, W.J.Xie. Calling patterns in human communication dynamics. Proc Natl Acad Sci, 110(5):1600-1605, 2013.
- [10].Xiaofan Wang, Xiang Li, and Guanyong Chen, Introdection to Network Science. Beijing: Higher education press,2012.