

Research on the Structure of Schoolmate Relationship Network*

Xiangjian Jia¹, Ning Zhang¹

¹Business School University of Shanghai for Science and Technology, Shanghai, China
jiaxiangjian@126.com, zhangning@usst.edu.cn

Abstract - A primary model of schoolmate relationship network is built in this paper based on social network services, topological structure of this network is studied combined with the knowledge of complex network theory. The actual relational data was obtained from an online social networking site-Renren.com open to college student, and the schoolmate network was constructed upon the database. Statistical quantities such as degree distribution, average shortest path length, clustering coefficient, assortativity coefficient etc. are calculated and analysed to capture the feature of the network. It is observed that the network is scale-free, has small-world property, and exhibits disassortative mixing pattern. Mechanism for the formation of some characteristics is also analysed and primarily illustrated, providing empirical foundation for further research about dynamical behaviours.

Index Terms - online social network; topological structure; complex network.

1. Introduction

Complex network is used to describe the relationship of interaction between individuals in complex systems like nature, society and engineering technology, which has been one of the hot issue studied by many subjects internationally since the end of 1980s [1-5]. So far, it has been extensively applied to mathematics, physics, biology, computer science, social economics and system science, and as the Social Network being a typical type of complex network, it has gaining more and more attention from researchers.

Everyone has an interpersonal relationship network in real life. In 1967, psychology professor Stanley Milgram from Harvard put forward the theory of 'six degrees of separation', which is the first empirical study on social network [6]. Social network analysis has become an important tool to explore individual behavior and community dynamics, and researching on social network has also practical significance.

Internet and WWW have dramatically changed people's mind and way of communication. Online Social Network rises fast and becomes mainstream, which draws people's attention and participation all over the world. Various kinds of social network sites have emerged nowadays; some famous sites are presented in TABLE I.

Online Social Network is obviously a special network, which has tremendous users and can be seen as a complex macro system. Many researchers domestic and overseas begin to study the complexity of SNS with theory of complex

network, and obtained some achievements. [8-10]

TABLE I Online Social Network Sites^[7]

Sites	Registered Users	Monthly Active Users	Date
Facebook	1000,000,000+	1000,000,000	2012.10
Twitter	500,000,000+	100,000,000	2012.4
Google+	400,000,000+	100,000,000+	2012.9
Sina Weibo	368,000,000	37,000,000	2012.6
Tencent Weibo	310,000,000	50,000,000	2011.12
Renren	160,000,000+	31,000,000+	2011.4

2. Statistical Indicators of Network Topology

A. Degree Distribution

Degree is the simple but very important node attribute; degree k_i of node i is defined as number of nodes that connect to node i . The average of all node degree defines the average degree, noted as $\langle k \rangle$. There are two ways to describe degree distribution:

1) Degree Distribution Function $P(k)$

2) Cumulative Degree Distribution Function P_k

$$P_k = \sum_{j \geq k} P(j). \quad (1)$$

Researches show that lots of social networks have power-law degree distribution [11], which can be expressed by formula (2).

$$P(k) \propto k^{-\gamma}. \quad (2)$$

It means that most nodes in social network have small degree; however there still exist few nodes have great degree. Formula (2) declares that social network is scale-free.

B. The Clustering Coefficient

The clustering coefficient is a local parameter of a network. Suppose node i has degree k_i , there actually exists number E_i of edges between these k_i neighbours, and the total number of edges can be $k_i(k_i - 1) / 2$, so we can define the clustering coefficient of node i as:

* This work is supported by the National Natural Science Foundation of China under Grant #70971089, Shanghai Leading Academic Discipline Project under Grant #XTKX2012, and the Innovation Fund Project For Graduate Student of Shanghai (JWCXSL1202).

$$C_i = \frac{2E_i}{k_i(k_i - 1)}. \quad (3)$$

Then the clustering coefficient for the whole network C is the average of all C_i . Researches proved that the values of clustering coefficient tend to be considerably higher than a random graph with a similar number of nodes and edges, which presents strong connectivity.

$C(k)$ stands for the average value of clustering coefficient of node with degree k , if it turns out $C(k) \propto k^{-\alpha}$, the networks should show some degree of hierarchy[14]. This is generally the feature of a nontrivial architecture in which small-degree nodes are well-clustered around the hubs and organized in a hierarchical manner into increasingly large groups [15].

C. The Average Shortest Path Length

The Average Path Length of network l is the mean of distance between any node pairs [4],

$$l = \frac{1}{1/2n(n+1)} \sum_{i \geq j} d_{ij}. \quad (4)$$

Where n is total number of node, d_{ij} is the geodesic distance from node i to node j . And the diameter of a network is defined as the maximum of the shortest path length, noted as $D = \max_{i,j} d_{ij}$.

In social network, l stands for the average number of friends on the shortest path connecting two individuals. Although many actual complex networks have huge number of nodes, research found that the average path length is remarkably short.

D. Degree Correlations

Another important element characterizing the local organization of complex networks is the degree. The Pearson correlation coefficient r is used to describe degree correlation coefficient degrees at either ends of an edge [16],

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}. \quad (5)$$

where j_i, k_i are the degrees of the nodes at the ends of the i th edge, with $i=1,2,\dots,M$, M is the total number of edges. The range is $-1 < r < 1$, the network is positive correlated if $r > 0$ and vice visa.

A large number of empirical studies [14] demonstrated that technical and biological networks are in general disassortative. But many online social networks turn out to show different pattern, which highlights the significant difference between online and actual social network.

E. Betweenness

There are two types of betweenness, node betweenness and edge betweenness [4]. And the node betweenness is generally used for many cases, which can be described as the number of shortest paths pass via this node. Suppose there are δ_{st} shortest

paths between node s and t , and $\delta_{st}(i)$ paths pass via node i , so we can define the betweenness of node i :

$$b_i = \sum_{s \in V} \sum_{t \neq s \in V} \frac{\delta_{st}(i)}{\delta_{st}}. \quad (6)$$

Betweenness can also be viewed as a measure of the ‘efficiency’ of the node, which reflects the hub level and influence power of node, indicating that many geodesic paths will get longer when this node is removed from the network.

3. The Structure of Schoolmate Relationship Network

A. Data Set

The actual relational data was obtained from Renren.com—a Chinese social networking site open to college students. The schoolmate network was constructed upon one user and all the relationship he had with his schoolmates. According to the rule of website, one applied to ‘add friend’ to another, they will become mutual friends if apply approved; so we got an undirected and unweighted network with 404 nodes and 8435 edges, as showed in Fig. 1.

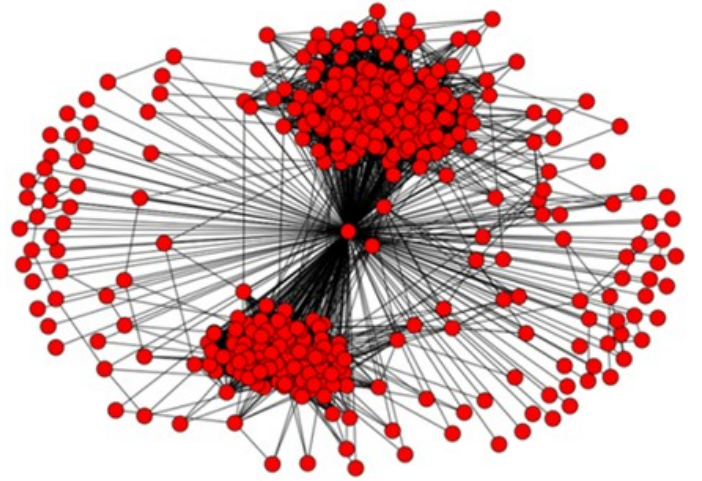


Fig. 1 Topology of Schoolmate Relationship Network.

B. The Analysis of Topological Structure

In this paper, Cumulative Degree Distribution P_k is applied to describe node degree distribution, as displayed in Fig. 2. Clearly, P_k of schoolmate network follows two different scaling of power-law, which can be seen as two sections divided by the specified critical degree value $k_c = 45$. P_k follows a power-law form $P(k) \propto k^{-\gamma_1}$ with $\gamma_1 = 0.25 \pm 0.01$, when $k < k_c = 45$, $P(k) \propto k^{-\gamma_2}$ with $\gamma_2 = 4.48 \pm 0.03$ for $k > k_c$. The average degree of network $\langle k \rangle$ is about 42, very close to the critical degree value.

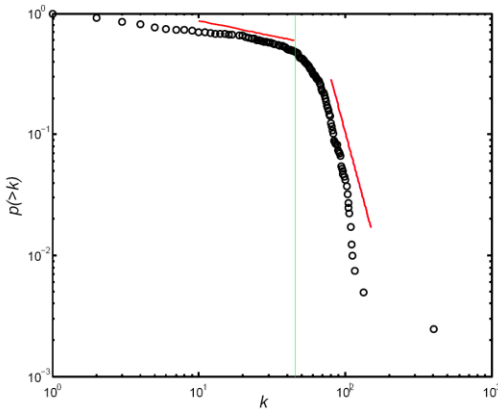


Fig. 2 Cumulative Degree Distribution of Schoolmate Network

The mathematics and physics meanings of two scaling are not mentioned in this paper, which can be found in reference [12-13].

In TABLE II, the node degree proportion distribution of this network is calculated. Left part of the table presents the top 10 degree values for its percentage ranking, and degree range $k \leq 5$ takes over the first five, indicating that most nodes of network have few connections. Right part shows the degree range proportion distribution and the weight of range 1~10 are as high as 30.4%, which is consistent with the left part. Percentage of node with high-degree is very small, which means in reality only few individuals have lots of connections or schoolmates.

TABLE II Degree Proportion of Schoolmate Network

k	Percentage (%)	[k]	Percentage (%)
1	7.4	1~10	30.4
2	6.9	11~20	5.4
3	4.7	21~30	6.9
4	4.2	31~40	6.7
5	2.5	41~50	8.2
48	2.2	51~60	10.9
68	2.0	61~70	9.4
9	1.7	71~80	10.4
79	1.7	81~90	4.4
20	1.5	91~100	3.5
		101~403	3.7

It is noteworthy that the medium node degrees of this network occupy certain proportion, and there are about 30% degree value belong to range of 51~80. This is because this network is composed by friends; most are schoolmates, of this particular user. According to the rule of Renren's mechanism, people are more likely to know each other in the special group of schoolmates. Therefore, high school classmates of this user are connecting more with each other, and come into being subgroup of high school; and the same as college classmates (in the same class or department), which form subgroup of college classmates. Two groups of nodes constitute two distinct communities separately, as illustrated in Fig. 4. Some nodes

link to both high school and college subgroup, which stands for these nodes are not only high school classmates of user but also come from the same university. Among the community, nodes connect to each other with high probability; as a result, the node degree will be relatively large. This can explain why there exists certain percentage of medium node degree.

Most nodes of the network have very small value of degree, which only have connects with user itself or few other individuals. These friends have little interaction with classmates in the network, because they may come from other school or get to know user in some website activities. Only a few nodes in the network have much larger degree, and the largest one is definitely the user. However, compared with actual situation, these nodes with large degree (except the user) are usually not the person who social widely in real life; but they performance much more 'active' in the online social network. The social website platforms provide 'social opportunity' for people who are 'inactive' in real life, making them become 'hubs' (large degree) with much more popularity. Maybe this is one of the driving forces that made online social networks increasingly rage.

The clustering coefficient of this schoolmate network is $C = 0.65$, order of magnitude much higher than that of a corresponding random graph of same size, indicating this network is strongly clustered. Fig. 3 plots the distribution of $C(k)$ vs. k ; a solid line is added for clarity. The fitted value of $C(k)$ is approximate to a constant (0.6), which is similar to $C = 0.65$. And the value of $C(k)$ diminishes with the increase of degree. Therefore, this network's clustering coefficient accords with the regular pattern, but cannot prove that obvious hierarchy exists in the network.

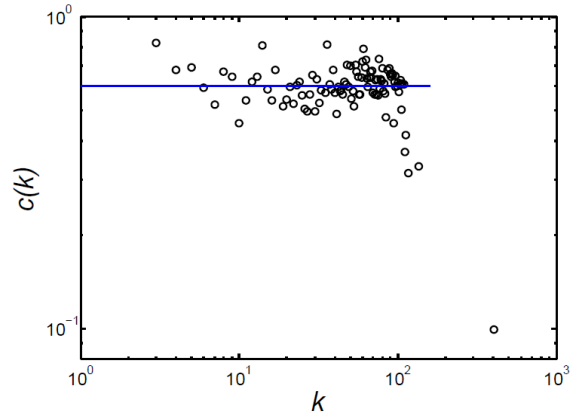


Fig. 3 Degree-dependent Clustering Coefficient $C(k)$ vs. Degree k

Average shortest path length is calculated, $\langle l \rangle = 1.98$ the same order of magnitude for logarithm of network scale ($\log 404 \approx 2.6$). As there is no isolated node in this connected network, well connectivity could decrease the shortest path length. Small average shortest path length $\langle l \rangle = 1.98$ and large clustering coefficient $C = 0.65$ make it clear that this observed network has small-world property.

Degree correlation coefficient is the same with many other online social networks, $r = -0.13$ presenting disassortative mixing pattern. In this network, nodes with small degree tend to link to that with large degree, which meaning that users with less friends are more likely to have connections with users who have lots friends. This is dramatically different with the situation of real life. Ordinary person expect to build relationship with celebrities or elites in real life, but these high-class person prefer to associate with person of same status. This invisible barrier of social class has been broken down by online social networks; ordinary individuals hence have chances to connection with the one with high popularity [17], so appears the disassortative mixing of degree.

The average betweenness value of this schoolmate network is 2.22×10^{-3} , the relationship of betweenness and degree is plotted in Fig. 4. It is clear that the betweenness increase gradually as the degree value become larger. This fitted with practical cases: the influence force of one person will be much more powerful when he 'knows' many people in his social circle.

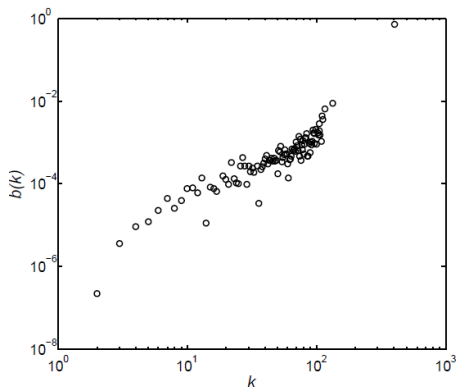


Fig. 4 the Relationship of Betweenness and Degree.

4. Conclusions and Outlook

Topological structure of one schoolmate network is analyzed in this paper. Large clustering coefficient and small average shortest path length make it clear that this observed network has small-world property. Degree distribution follows two different scaling of power-law, indicating that this network is scale-free. Since degree correlation coefficient is negative, this schoolmate network presents disassortative mixing pattern. The clustering coefficient approximates to a constant, which means there is no hierarchy in this network.

As online social networks are increasingly popular and gradually become a necessary part of our daily life, so there are potential theory and application values researching the online social network. This type of networks is constructed upon common interests, common business background and trust; on the contrary, the connections of network can also affect our interests and our trust between each other. Profound understanding of topological structure of online social network can contribute to realizing the spread progress of information on the network, to knowing the formation or strengthen or remove of the trust among individuals, to optimizing the traffic

flow more reasonably [18]. The research of online social network also provides unprecedented opportunities for socialists and psychologists to probe into social network of huge scale, which may assist them to find new type of collective behaviors and human dynamic behaviors. Political election can also be influenced by online social network at some degree [19]; meanwhile, salesman could use this type of network to promote their commodity more efficient.

Except the structural characteristic, there are more directions on researching online social network: analysis of user behavior, social recommendation based on social trust, community mining, information spreading, innovation diffusion, and network evolution etc.

5. References

- [1] D-J. Watts, and H-Z. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, June 1998.
- [2] R. Albert, H. Jeong, A-L. Barabási, "Diameter of the World Wide Web," *Nature*, vol 401, pp. 130-131, September 1999.
- [3] A-L. Barabási, R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509-512, October 1999.
- [4] W. Xiaofan, L. Xiang, C. Guanrong, "Complex network theory and applications," Beijing: Tsinghua University Press, 2006.
- [5] Z. Tao, B. Wenjie et al, "A brief review of complex networks," *Physics*, vol. 34, no. 1, pp. 31-36, 2005.
- [6] S. Milgram, "The small-world problem," *Psychology Today*, vol. 1, no. 1, pp. 61-67, May 1967.
- [7] Wikipedia, "List of social networking websites," 2012. http://en.wikipedia.org/wiki/List_of_social_networking_websites.
- [8] J. Bollen, B. Goncalves, G-C. Ruan et al, "Happiness is assortative in online social networks," *Artificial Life*, vol. 17, no. 3, pp. 237-251, May 2011.
- [9] K. Lewis, M. Gonzalez, J. Kaufman, "Social selection and peer influence in an online social network," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 1, pp. 68-72, December 2011.
- [10] Z. Yanchao, L. Yun, Z. Haifeng et al, "The research of information dissemination model on online social network," *Acta Phys. Sin.* vol. 60, no. 5, pp. 1-7, May 2011.
- [11] M-E-J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167-256, March 2003.
- [12] W-J. Reed, "The Pareto law of incomes: an explanation and an extension," *Physica A*, vol. 319, pp. 469-486, 2003.
- [13] W-J. Reed, M. Jorgensen, "The double Pareto-lognormal distribution: a new parametric model for size distribution," *Stats-Theory & Methods*, vol. 33, no. 8, pp. 1733-1753, 2004.
- [14] E. Ravasz, A-L. Barabási, "Hierarchical organization in complex networks," *Phys Rev E*, vol. 67, no. 2, pp. 026112, 2003.
- [15] H. Haibo, W. Ke, X. Ling et al, "Analysis of online social networks based on complex network theory," *Complex system and complexity science*, vol. 5, no. 2, pp. 1-14, 2008.
- [16] M-E-J. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, no. 20, pp. 28701, 2002.
- [17] L. Yongjun, "Analysis on topological features of online social networks," *Complex system and complexity science*, vol. 9, no. 3, pp. 22-37, 2012.
- [18] A. Mislove, M. Marcon, P-G. Krishna et al, "Measurement and analysis of online social network," *Proc. 7th ACM SIGCOMM Conf.*, pp. 29-42, 2007.
- [19] Phys.org, "Obama and Romney's social media face-off," 2012. <http://phys.org/news/2012-10-obama-romney-social-media-face-off.html>.