Study of the Growth of a New Social Network Platform

Ning Tang
Mobile Life and New Media Laboratory
Beijing University of Posts and Telecommunications
Beijing, China
ning.tang0@gmail.com

Chunhong Zhang
Mobile Life and New Media Laboratory
Beijing University of Posts and Telecommunications
Beijing, China
zhangch.bupt.001@gmail.com

***Abstract*—Nowadays, with the rapid development of social network sites (SNS), many efforts have been made on the analysis of such networks, which is important for us to enrich our social network knowledge and improve the SNS design. However, most analysis have focused on the very popular SNS, there is little comprehensive analysis of growing process of a SNS from the exact original stage, therefore, lacking of deep understanding how the SNS evolve over time. In this paper, a comprehensive growth data set of a new social network site constructed by our lab is examined to see the growing process of the network and to find some underlying growth mechanisms. This observation provides an empirical evidence of the Barabási-Albert (BA) model. During the observation, the number of new added links is shown to have a linear relationship with the number of new added users and there is a preferential attachment phenomenon in the link formation process. In addition, the degree distribution at different time points is presented to see the formation of the power-law distribution. Finally, the topological properties are found to be stable after a period of development, though the network is still growing.**

Keywords-social network sites; growth; power-law degree distribution; topological properties.

I. INTRODUCTION

There are various social network sites nowadays. On SNS, users can make friends, share pictures, share videos, write blogs and play games. Despite the different goals and purposes of various social network sites, they have been shown to have a number of common structural features, such as power-law degree distribution, small diameter, and high clustering coefficient [1][2]. Many previous analysis have focused on validating static common features in different popular social network sites, however, there is little comprehensive analysis of the growing process of these features from the exact original stage of a SNS, therefore, lacking of deep understanding how the features form over time. Although there are some theoretical models [3] that try to reveal the underlying mechanisms, an empirical view of the formation process is desired.

An in-depth understanding of the growing process of a SNS from the exact original stage can help us learn the underlying mechanisms which result in the specific features. Mastering the mechanisms enables us to simulate similar social networks and provides a possibility for further social networking study. What is more, a better knowledge of the evolution of the topological properties allows us to predict the future of networks and make corresponding improvements.

In this paper, an empirical analysis of the growing process of a new social network site from the exact original stage is presented to see how its features shape over time. The social network is constructed by our lab. The dataset used in this paper is the testing data extracted from the server for 27 consecutive days. Compared with the crawled data sets generally used in previous studies [2][4][5][6], the data set used in this paper is a complete and time continuous one which may help us make a more comprehensive and more accurate understanding of the whole network.

The number of new added links is found to have a linear relationship with the number of new added users—every 11 additional edges will bring about 6 new users. And there is a preferential attachment phenomenon in the link formation process—more than 55% of new links are attached to the top 30% large-degree nodes. Next, the degree distribution at different time points is plotted to see the dynamic shaping of the power-law distribution. Finally, the topological properties are found to be stable after a period of development, though the network is still growing.

The rest of this paper is organized as follows: Additional background and related works are provided in Section 2. In Section 3, the methodology for obtaining the data set and its limitations are described. In Section 4, an empirical analysis of the growing process of the new social network is presented. Finally, the conclusion remarks and future work are drawn in Section 5.

II. BACKGROUND AND RELATED WORK

A. Topological Properties

There are some general metrics to characterize the SNS, which are called topological properties.

- Degree. Degree is the number of edges incident to a vertex of a graph. The node degree distributions of many large-scale social networks have been shown to conform to power-laws. Power-law networks, also known as scale-free networks, are networks where the probability that a node has degree k is proportional to $k^{-\gamma}$, for large k and $\gamma > 1$. The parameter γ, whose value is typically in the range $2 < \gamma < 3$, is called the power-law coefficient. In power-law network, the majority of the nodes have small degrees, but a few nodes called hubs have significantly high degrees. In 1998, Barabási and Albert found that growth and preferential attachment were two important mechanisms of the formation of power-law networks [3]. Growth means that the number of nodes in the network increases over time. Preferential attachment means that new links are more likely to attach to large-degree nodes. They later proposed the Barabási and Albert (BA) model based on the two mechanisms.

- Diameter. The shortest path length L between two vertices in a graph is the number of edges in a shortest path connecting them. The average shortest path length <L> is the mean of L between any pairs that have at least a path connecting them. Diameter D is defined as the maximum of the shortest path length.
- Clustering coefficient. Clustering coefficient is a measure of degree to which nodes in a graph tend to cluster together. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups with a relatively high density of ties [7][8]. Clustering coefficient of node i in undirected network is defined as

$$C_i = 2E/\, K(K-1). \tag{1}$$

That is the ratio of the number E of edges that actually exit between the k neighbors of node i to the potential number $k(k-1)/2$. The clustering coefficient C of the whole network is the average of all individual C_i.

Studies have shown that the Web [9][10], scientific collaboration on research papers [11], film actors [12], and general social networks [13] have small-world properties. Small-world networks have a small diameter and exhibit high clustering.

B. Related Works

Recently, much work has focused on understanding the structure and evolution of large-scale online social networks. [2][4], present empirical analysis of statistical properties and validate the power-law, small-world and scale-free properties of several large-scale social networks. Bimal [5] finds that while the individual links that constitute the activity network change rapidly over time, the average network properties remained relatively stable in Facebook. Alan [6] shows that new link formation in Flickr follows preferential attachment, but the link creation process cannot be explained by the BA model alone because users are far more likely to link to nearby users than that model would suggest.

Above researches are all about popular social networks which have already processed the common features. But the tracing observation of the formation of a new SNS, which is helpful for a better understanding of the underlying growth mechanisms and helpful for predicting the future of networks, is lack of studying.

III. Methodology

To get the growth data of a SNS from the exact original stage, a new SNS is constructed by our lab. Then the new SNS began to get a test from September 15th, 2010. The dataset used in this paper is the user data extracted from the server at the end of the test on October 11th. During the 27 days, 140 users have registered on the platform, including 110 undergraduates in four classes at Beijing University of Posts and Telecommunications and 30 graduate students in Mobile Life and New Media Laboratory. The dataset contains all the link formation information of the 27 days, including the creator, the target and the timestamp.

In contrast to the crawled data sets used in other papers [2][4][5][6], the data set used in this paper is a complete and time continuous one, which may help us make a more comprehensive and accurate understanding of the whole network. However, although our dataset contains all the user data, the statistical properties may not be so obvious because the number of testing users is small and the testing period is short.

IV. Data Analysis

The network is composed of users (vertices) and links (edges) among them. Since link creation in our network requires consent from the link target, a link connecting the creator and the target is undirected. Among the 140 registered users, there are 47 users who have at least one link with others, while the remaining 93 people are isolated nodes. In the next analysis, we'll only consider users that connect with others.

In order to learn the evolution of the network, the growing process of the network is divided into segments. During the test, registration is mainly concentrated in a few days and the number of registers varies greatly every day (from 0 to 49 per day). To make the partition as even as possible, the growing process is divided by the number of new added links instead of by days. Finally, the network developed to have 77 edges. Considering granularity, the growing process is finally divided into seven segments. In each segment, 11 edges are added into the network. In the following analysis, the network is examined at each end of the seven segments. Table Ⅰ shows the number of vertices and edges at the seven time points.

TABLE I. THE NUMBER OF VERTICES AND EDGES AT SEVEN TIME POINTS

Time point	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Edges	11	22	33	44	55	66	77
Vertices	10	15	22	30	36	41	47

 ISBN: 978-1-61208-161-8

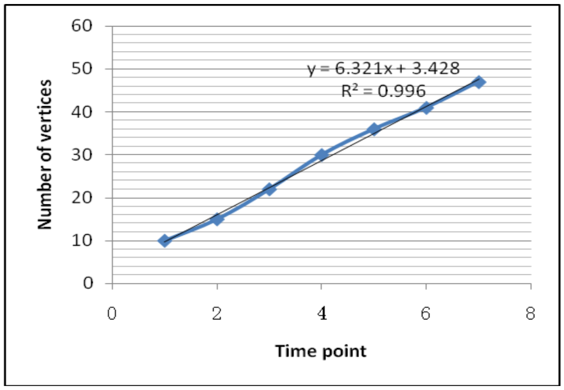


Figure 1. Plot of the number of vertices at seven time points

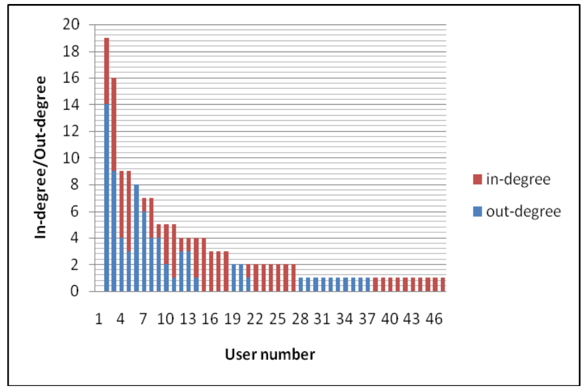


Figure 2. The statistics of the in-degree and out-degree of each user

A. Preferential Attachment Phenomenon

The well-known Barabási-Albert (BA) model has been shown to result in networks with power-law degree distributions. In BA model, new links are attached to nodes using a probability distribution weighted by node degree. Since the dataset used in this paper is very small, there could not be a statistic of the link formation distribution with degree. Instead, the number of links that are attached to nodes whose degrees rank in the top 30% at the former time point is calculated to examine whether there is a preferential attachment phenomenon in the link formation process. The result is shown in Table Ⅱ.

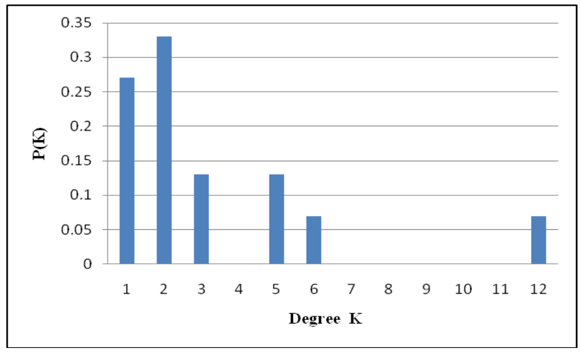
It is obvious that there is a preferential attachment phenomenon—more than 55% (6/11) of new links are attached to the top 30% large-degree nodes.

Since the establishment of a link needs a creator and a target, so the link can be regarded as a directed one in this respect. Here we define the out-degree of a node as the number of links that it creates, and the in-degree as the number of links that it receives. Figure 2 is the statistics of the in-degree and out-degree of each user, ranking in descending order of the total degree. It is obvious that large-degree nodes are more likely to be creators.

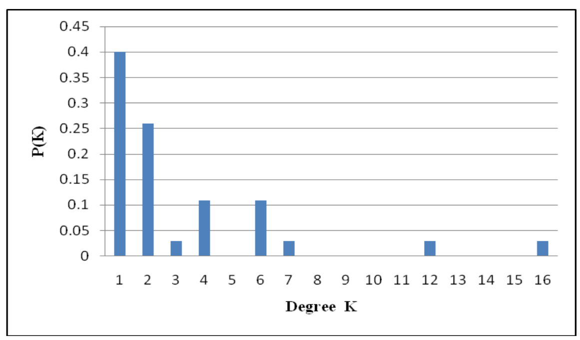
From Table Ⅱ and Figure 2, the underlying mechanism which results in the power-law degree distribution can be figured out. The large-degree nodes are generally active users who like to create links, so new links are more likely to be established between large-degree nodes and new added vertices. What is more, since every 11 additional edges generally bring about 6 new users, new nodes are continuously added to the network, making the degree distribution more and more asymmetrical, as shown in Figure 3—more and more nodes have small degrees, while a few hub nodes have larger and larger degrees.

TABLE II. THE NUMBER OF LINKS THAT ARE ATTACHED TO THE TOP 30% LARGE-DEGREE NODES

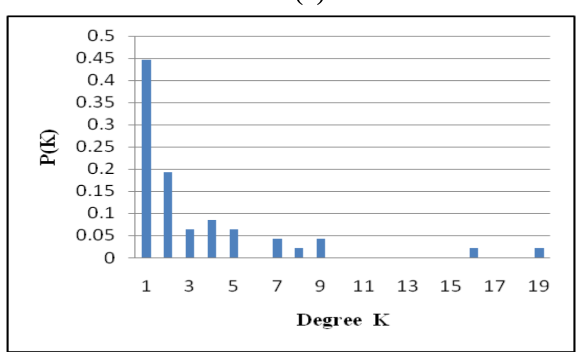
Time point	1	2	3	4	5	6	7
Edges	-	7	8	6	9	6	6



(a)



(b)



(c)

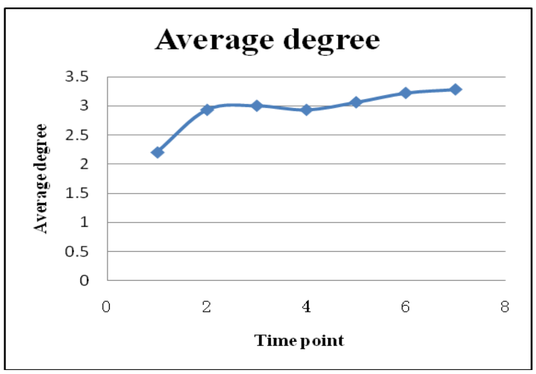
Figure 3. The change of degree distribution at three time points.

B. *Topological Properties*

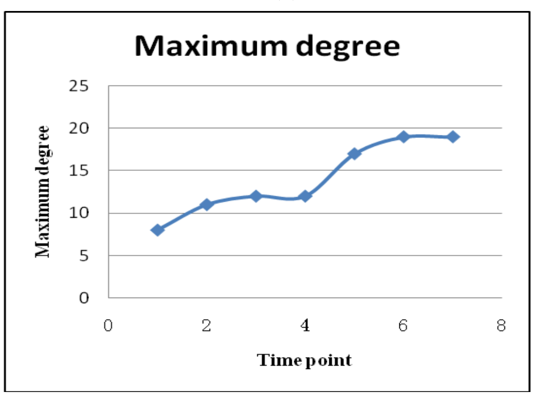
Next there will be a look at the evolution of the topological properties. Table III is the detailed growing information at seven time points. Figure 4 plots the changing processes of these properties. It is shown that although the network keeps a same expanding speed during each time point, all the properties tend to be stable from the fifth time point. The network seems to have entered a stable stage.

TABLE III. THE TOPOLOGICAL PROPERTIES AT SEVEN TIME POINTS

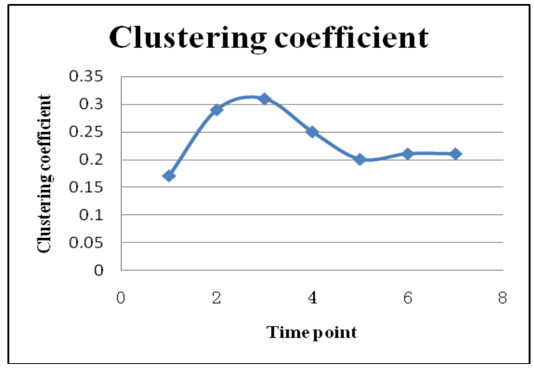
Time point	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Average degree	2.2	2.93	3	2.93	3.06	3.22	3.28
Maximum degree	8	11	12	12	17	19	19
Clustering coefficient	0.17	0.29	0.31	0.25	0.2	0.21	0.21
Average shortest path length	1.89	2.06	2.24	2.39	2.73	2.72	2.77
Diameter	3	4	4	4	6	6	6



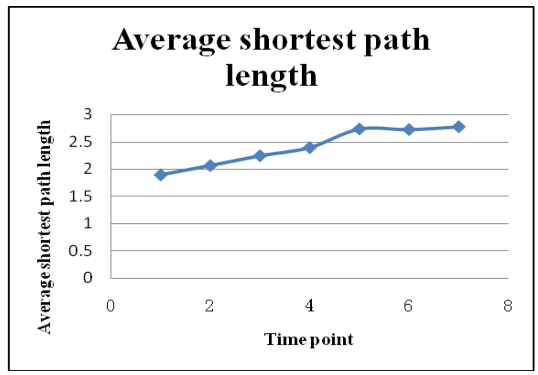
(a)



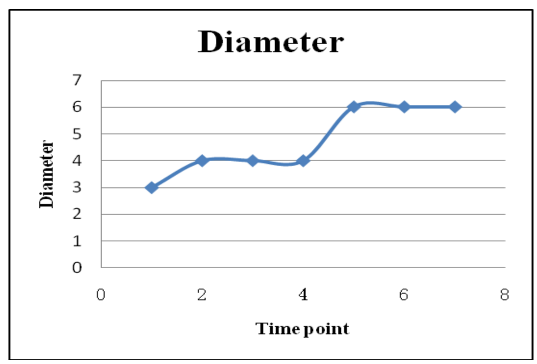
(b)



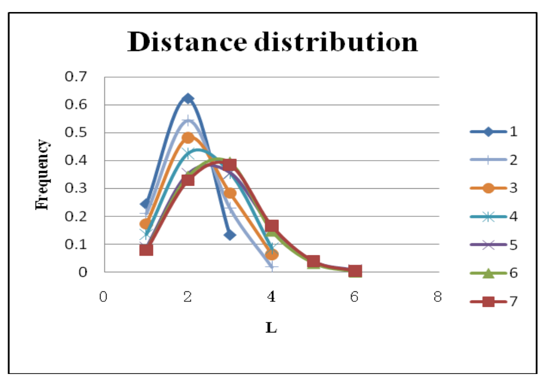
(c)



(d)



(e)



(f)

Figure 4. The evolution of the topological properties

In order to affirm the new developed network used in this paper has the common features that previously observed in other large-scale social networks and ensure the above study is applicable to general social networks, there is a check of the power-law and small-scale properties in the new network. Figure 5 is the degree distribution of the final network.

In Figure 5, majority nodes have small degrees (k=1, 2), some nodes have moderate degrees (k=3~9), few nodes (the long tail) have very large degrees (k=16, 19). It presents an approximate power-law distribution. To a power-law network, the cumulative degree distribution also follows a power-law form as $P(\geq K) \propto k^{-(\gamma-1)}$. The above function plotted in log-log coordinates is a straight line with slope $-(\gamma—1)$. Figure 6 is the log-log plot of the cumulative degree distribution. Using linear least squares regression to get the slope $-(\gamma—1) = -1.268$. So the degree distribution follows a power-law form with γ=2.268 and the network is a scale-free network.

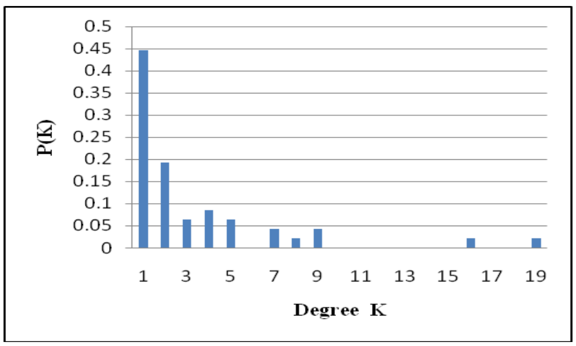


Figure 5. Degree distribution

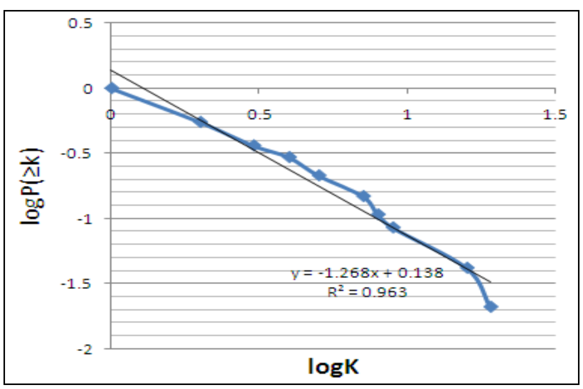


Figure 6. Log-log plot of the cumulative degree distribution

The clustering coefficient C of the whole network is 0.21, much higher than that of a corresponding random graph of the same size C_{rand}=<k>/N=3.28/47=0.07, <k> is the average degree of the undirected network. In addition, the network also has a small average shortest path length 2.77. Hence, small-word property also exists in this network.

At the end of the test, the social network has exhibited the small-world and scale-free properties after a short-term development. Therefore, social networks are born to have the common features. It also indicates that the above study of the new social network in this paper is applicable to general social networks.

V. CONCLUSION AND FUTURE WORK

In conclusion, the experiment in this paper provides an empirical evidence of the BA model. Both growth and preferential attachment mechanisms are observed in our observation. In addition, a new phenomenon that the number of new added users has a linear relationship with the number of new added links is observed. The topological properties are found to be stable after a period of development. Finally, the common small-world and scale-free properties are proved to exist in this small-scale social network, indicating that the study of the network is applicable to general social networks.

We think our work provides an insight into the understanding of the original stage of a social network, which will help us know the underlying mechanisms and predict the future growth. Much work remains to be done. We'll make a longer and larger-scale observation of the growing process of the network to get more statistical properties. In addition, other aspects of the network, such as user behaviors, are worth studying.

ACKNOWLEDGMENT

This work is supported by China-Finland Cooperation Project on the Development and Demonstration of Intelligent Design Platform Driven by Living Lab Methodology (2010DFA12780), Architecture and key Technology Research on Broadband Wireless Campus Innovation Testbed (2010ZX03005-003), the National Key Program of New Generation of Broadband Wireless Mobile Communication Network (2011ZX03005-004-02) and Key Laboratory of Universal Wireless Communications, Ministry of Education.

REFERENCES

[1] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of Topological Characteristics of Huge Online Social Networking Services," In Proceedings of the 16th World Wide Web Conference (WWW'07), Banff, Canada, 2007, pp. 835–844.

[2] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," In Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC'07), San Diego, CA, 2007.

[3] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," Science, vol. 286, 1999, pp. 509–512.

[4] Feng Fu, Lianghuan Liu, and Long Wang, "Empirical analysis of online social networks in the age of Web 2.0," Physica, vol. 387, 2008, pp. 675–684.

[5] Bimal Viswanath, Alan Mislove, and Meeyoung Cha, "On the Evolution of User Interaction in Facebook," WOSN'09, August 17, 2009.

[6] Alan Mislove, Hema Swetha, and Koppula Krishna P . Gummadi, "Growth of the Flickr Social Network," WOSN'08, August 18, 2008.

 ISBN: 978-1-61208-161-8

[7] P. W. Holland and S. Leinhardt, "Transitivity in structural models of small groups," Comparative Group Studies, vol. 2, 1998, pp. 107–124.

[8] D. J. Watts and Steven Strogatz, "Collective dynamics of 'small-world' networks," Nature, vol. 393 (6684), June 1998, pp. 440–442.

[9] R. Albert, H. Jeong, and A.-L. Barabási, "The Diameter of the World Wide Web," Nature, vol. 401, 1999, p. 130.

[10] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph Structure in the Web: Experiments and Models," In Proceedings of the 9th International World Wide Web Conference (WWW'00), Amsterdam, May 2000.

[11] M. E. J. Newman, "The structure of scientific collaboration networks," Proceedings of the National Academy of Sciences (PNAS),vol. 98, 2001, pp. 409–415.

[12] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," Proceedings of the National Academy of Sciences (PNAS), vol. 97, 2000, pp. 11149–11152.

[13] L. A. Adamic, O. Buyukkokten, and E. Adar, "A social network caught in the Web," First Monday, 8(6), 2003.