Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A last updating evolution model for online social networks



^a College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China ^b Computer and Information Sciences, The University of Alabama at Birmingham, USA

ARTICLE INFO

Article history: Received 19 April 2012 Available online 17 January 2013

Keywords: Online social network Evolution model Last updating time Scale-free Small-world effect

ABSTRACT

As information technology has advanced, people are turning to electronic media more frequently for communication, and social relationships are increasingly found on online channels. However, there is very limited knowledge about the actual evolution of the online social networks. In this paper, we propose and study a novel evolution network model with the new concept of "last updating time", which exists in many real-life online social networks. The last updating evolution network model can maintain the robustness of scale-free networks and can improve the network reliance against intentional attacks. What is more, we also found that it has the "small-world effect", which is the inherent property of most social networks. Simulation experiment based on this model show that the results and the real-life data are consistent, which means that our model is valid.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

With the advance of information technology as well as the booming of the Internet, people are turning to electronic media more frequently for communication, and social relationships are increasingly found on online channels. Various interactive websites, such as BBS, Micro-blog, provide opportunities to everyone to discuss with others on all kinds of topics and spread the topics on the Internet [1,2]. People with common interests communicate with each other on BBS; their topic of interest is usually identified by the board itself. IDs registered by users are actors on BBS, through which one can review messages left by others and leave one's own messages if one wants, thereby forming a thread. Thus, a thread in the BBS roughly represents a dialog between people, and such a dialog constitutes the basic relationship among the people participating in it.

In online social networks, dialogs or discussions usually proceed with little restriction on message writing and discrimination based on personal information, thereby forming the intense discussions. Therefore, the pattern of such online social relationships may be different from that of traditional social relationships based on face-to-face contact or online communication involving an exchange of personal information, such as e-mail transactions and instant messaging. On the basis of these observations, we think it is important to study the evolution model of online social networks constructed by people in intense discussions; this would be useful in resolving diverse sociological and political issues and understanding the manner in which social opinion is formed in the digital era.

The first time when mathematicians tried to describe a network might be traced back to 1736, when they were dealing with the famous "Konigsberg seven bridges" problem [3]. Then, Erdos and Renyi built their powerful random graph theory to model the random-like complexity of various networks in 1960 [4,5]. Motivated by this classical random graph theory, some new network models were proposed. The Watts–Strogatz small-world model depicts a large-clustering and small-average-path phenomenon in real world [6,7]. Barabasi and Albert proposed a scale-free model to explain the "rich gets





CrossMark



STATISTICAL MICHANI

PHYSICA

^{*} Corresponding author at: College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. *E-mail addresses:* buzhan@nuaa.edu.cn, zhanb@cis.uab.edu (Z. Bu), zhengyou_xia@nuaa.edu.cn (Z. Xia), aics@nuaa.edu.cn (J. Wang), zhang@cis.uab.edu (C. Zhang).

^{0378-4371/\$ –} see front matter 0 2013 Elsevier B.V. All rights reserved. doi:10.1016/j.physa.2013.01.006

richer" phenomenon, in which degree distribution obeys power-law $P(k) \sim k^{-3}$ [8,9]. The competition aspect was subsequently discussed in their fitness model for the fitness-gets-richer phenomenon, and they pointed out that a generalized power-law or stretched exponential degree distribution follows the fitness selection [10]. Recently, a local-world evolving network model has been proposed by Xiang Lia and Guanrong Chen, which considers the distance preference [11].

Although many evolving network models have been used to analyze possible hidden relationships under specific evolving mechanisms, we have noticed some other important factors that had been ignored by previous researches [4–12]. In this paper, we will focus on a real-life online social network-Tianya (http://www.tianya.cn), which is one of the largest online social networks in China at present. It is observed that, the global preferential attachment mechanism does not work for those users (represented by nodes in the graph) that have less than 100 comment connections with other users. In other words. Tianya community is organized by some continuous active users who can be called forum moderators or super fans. Their activity is manifested in two aspects: (1) uninterruptedly post topics; (2) initiatively comment others. Assume that the network is dynamically growing; the new coming nodes are more likely interact with those active ones. On the other hand, users who have just participated in the discussions on the Internet are more likely to acquire attentions from others. This property has been confirmed in many online social networks, such as DigitalPoint, BlackHat, and ABestWeb. On the basis of these observations, the preferential attachment mechanism does not work on the global network, but does work on the "last updating time" of each node. Here, the concept of "last updating time" can represent a characteristic of online users, which means the time interval from his last interaction on Forum to present. Therefore, establishing and studying a last updating network model will enable us to better understand and describe more real-life online social networks. Two most natural questions are: how to build such a model and how much effect does a last updating model have on its network's scaling exponent, scale-free property, and complex dynamics? This paper attempts to provide some answers to these questions.

The remainder of this paper is organized as follows. In Section 2, we provide an introduction to the research motivation and a description of the study data. The new evolution model based on "last updating time" is introduced and studied in Section 3, followed by some discussions on computer simulations in Section 4. Finally, Section 5 concludes the investigation.

2. Dataset and motivation

In this section, the data we used and the research motivation are introduced.

2.1. Data and preprocessing

The data used in this paper were downloaded from Tianya.com, a popular bulletin-board service in China. It includes more than 300 boards, and the total number of registered user identifications (IDs) is more than 32 million. Since its introduction in 1999, it has become the leading social-networking site in China due to its openness and freedom. Each article on Tianya contains the author ID (the user ID posting the current article), title, board information, date and time, and contents; if the post is a reply article, the replier ID (the user ID who posted the article that the current article comments on) and replied contents are also included. All the above information is regularly distributed in an HTML source file; here, we implemented a relatively simple analysis tool to extract the data using regular expressions. Fig. 1 shows a typical HTML source file on Tianya.com, in which all the necessary components are marked.

To study the Tianya social network, we began by adopting the formalism in Ref. [13]. Every registered user identification (ID) corresponds to a node $i \in V$ in a graph $G = \langle V, E \rangle$. An edge $(i, j) \in E$ represents a social relation between two users that results from their comment activity. Let n_{ij} be the number of times that user i writes a comment to user j. Then, an undirected edge exists between users i and j if either $n_{ij} > 0$ or $n_{ji} > 0$.

In this study, we selected the worldview board on Tianya.com from which to collect statistics for the online social networks. The networks were created from the articles posted between July, 2003 and January, 2010, including 324,666 users, 99,735 threads and 4712,859 replies.

2.2. Motivation

The social structure of real-life online social networks is of interest to many researchers. A key aspect of studying these networks is to understand the evolutionary dynamics and the mechanism by which these structures grow and change over time. Fig. 2 shows the evolutionary dynamics of the worldview board on Tianya.com from 2002 to 2010. Some previous statistical analysis reveals that most real-life online social networks are consistent in nature (both the "small-world effect" [14] and skewed degree distributions [15,16] are found in them). However, traditional network evolution models cannot simultaneously depict those two characteristic. Although the random graph has small average shortest path length, it do not have obvious clustering characteristic; what is more, its degree distribution P(k) obeys Poisson distribution. The Watts–Strogatz small-world model depicts a large-clustering and small-average-path phenomenon; however, its degree distribution P(k) obeys uniform distribution. Above two models are only suited to describe static features of networks, they are less capable of specifying models of change and simulating network evolution. The Barabasi–Albert scale-free model can be used to simulate network evolution, and the degree distribution obeys power-law $P(k) \sim k^{-3}$; unfortunately, it inherits unobvious clustering characteristic of the random graph. The local-world evolving network model represents a



Fig. 1. A typical HTML source file with necessary components marked.



Fig. 2. The evolutionary dynamics of the worldview board on Tianya.com from 2002 to 2010.

transition between power-law and exponential scaling; still, it cannot depict a large-clustering network. On the basis of these observations, we hope to detect the real-life evolving mechanisms, with which we can model online social networks. Here, we present a novel evolution network model with the new concept of "last updating time".

3. The last updating evolution model

It is observed that, the real-life online social network is organized by some continuous active users. Their activity is manifested in two aspects: (1) uninterruptedly post topics; (2) initiatively comment others. Fig. 3 gives a simple graph model of Tianya community, in which the blue nodes constitute backbone networks. It is obvious that the average degree of those blue nodes (can be seen as active users) is relatively high. Assume that the network is dynamically growing; the new coming nodes are more likely interact with those active ones. On the other hand, users who have just participated in the discussions on the Internet are more likely to acquire attentions from others. As shown in Fig. 4, theme sequence on the worldview board is arranged according to topic update date. The similar property has been confirmed in other online social network, such as DigitalPoint (http://forums.digitalpoint.com), BlackHat (http://www.blackhatworld.com), and ABestWeb (http://forum.abestweb.com). Therefore, the preferential attachment mechanism does not work on the global network, but



Fig. 3. A section of real-life online social network (643 nodes and 650 edges).

does work on the "last updating time" of each node. Here, the concept of "last updating time" can represent a characteristic of online users, which means the time interval from his last interaction on forum to present. To model such "last updating time" effect, a last updating evolution model based on the local-world evolving algorithm is now proposed, to be generated by the following algorithm:

- (1) Growth: Start with a small number m_0 of nodes and small number e_0 of edges. At every time step, add a new node with $m(m \le m_0)$ edges that link the new node to *m* different nodes already present in the network. And each new node is assigned a "last updating time" parameter *Updated*_i, which is the time interval from his last interaction on forum to present (In initial condition, *Updated*_i = 0).
- (2) Local world: Select nodes whose "last updating time" is below the threshold ϕ ($0 \le \phi \le t$) from the existing network.
- (3) Preferential attachment: The probability $\prod_{Local}(k_i)$ that a new node is connected to node *i* depends on the degree k_i of node *i*, in such a way that:

$$\prod_{Local}(k_i) = \prod'(i \in LW) \cdot \frac{k_i}{\sum_{j \ Local}k_j} \equiv \frac{\sum_{i} fun_i}{m_0 + t} \cdot \frac{k_i}{\sum_{j \ Local}k_j}.$$
(1)

In which, *fun_i* is the judging function, it is defined as:

$$fun_i = \begin{cases} 1 & \text{if } Updated_i \le \phi \\ 0 & else \end{cases}$$
(2)

 $\prod'(i \in LW)$ is the probability that node *i* is selected into the local-world network. "*LW*" refers to all the nodes in interest with respect to the new coming node at time step *t*.

(4) Updating: The "last updating time" of nodes in the end of new *m* edges is updated to 0, while for other nodes, $Updated_i = Updated_i + 1$.

Thus, in at every time step t, the newly coming node connects to m nodes, which are selected from its local world with preferential attachment. The local world is selected from existing nodes whose "last updating time" is below the threshold ϕ , but is not selected randomly in Xiang Lia and Guanrong Chen's evolution model. The "last updating time" of every node will be updated, depending on their actual interactive behavior. Different network topologies can be acquired by dynamically changing the threshold ϕ . We can study their corresponding statistical characteristics so as to compare them with real-life online social networks. In the rest of the paper, we explain our model in detail.

Z. Bu et al. / Physica A 392 (2013) 2240-2247



Fig. 4. The snapshot of the worldview board on Tianya.com.

Table 1

Statistics comparison of real-life online social network and the last updating evolving networks with threshold $\phi = 10$; 100; 1000; 10000, $m_0 = 15$, $e_0 = 30$, m = 10, respectively.

	Real net	Net with $\phi = 10$	Net with $\phi = 100$	Net with $\phi = 1000$	Net with $\phi = 10000$
Ν	323,745	300,015	300,015	300,015	300,015
М	2987,953	3000,030	3000,030	3000,030	3000,030
$\langle k \rangle$	18.46	20.00	20.00	20.00	20.00
С	0.0712	0.1021	0.0654	0.0124	0.0042
C_{rand}	0.000057	0.000067	0.000067	0.000067	0.000067
l	3.7781	3.9612	4.6111	4.2358	4.5254
lrand	4.3517	4.2099	4.2099	4.2099	4.2099
D	10	7	9	10	11

4. Computer simulations

In this section, the statistical properties of four artificial networks are analyzed, and we compare the results with the real-life online social network to characterize how they differ or resemble one another.

4.1. Global properties

Our statistical analysis revealed that the last updating networks exhibit several nontrivial topological properties, such as the "small-world effect" and skewed degree distributions. Here, we discuss network characteristics from a global perspective. Table 1 shows the statistics describing the five networks.

The clustering coefficient of a node *i* is defined as $C_i = \frac{2E_i}{k_i(k_i-1)}$, which is the ratio of the number of edges E_i that interconnect the k_i neighbor nodes of node *i* and their total possible number $k_i(k_i - 1)/2$; for $k_i = 1$, $C_i = 0$. The clustering coefficient of the whole network is the average of the individual C'_i [17]. We see that, for the first three graphs, *C* is much higher than the randomized counterpart C_{rand} [14]. The average shortest path length *l*[18] was computed; this path length is



Fig. 5. Variations with the threshold ϕ .

the mean of the geodesic distance between pairs of nodes connected by at least one path. This average shortest path length is small for all five networks. Moreover, this statistic is roughly the same as that for a random graph l_{rand} [14]. The diameter *D* of those networks, which is defined as the maximum of the shortest path length, is also very small. Large *C* values indicate that discussions can be begun among bunches of users easily. Small *l* values indicate that ideas and opinions can propagate rapidly from one person to another. Hence, the small-world topologies of online social networks ensure the propagation of discussions among users. From Fig. 5, we can conclude that if the threshold ϕ is set in a suitable range, for example, $10 \le \phi \le 100$, the networks we generated have "small world effect".

4.2. Degree distribution

The most basic topological characterization of a graph *G* can be expressed in terms of its degree distribution P(k) [19], which is defined as the probability that a node chosen uniformly at random has degree *k* or, equivalently, as the fraction of nodes in the graph having degree *k*. The analysis of this degree distribution describes the level of interaction between users and provides a robust indicator of the degree of heterogeneity within the network. In this section, we discuss the degree distributions of the four last updating networks.

It is obvious that at every time step t; $0 \le \phi \le t$, and there are two limiting cases in the above-proposed last updating evolving network model: $\phi = 0$ and $\phi = t$, which are further discussed below.

Case A: $\phi = 0$.

In this limiting case, the preferential attachment selection is not effective in the network growing process. This is the same as the case of model A in the Barabasi and Albert scale-free model [8,9], which keeps the growing manner without preferential attachment. The rate of change of the connectivity of vertex *i* in this case is given by

$$\frac{\partial k_i}{\partial t} = \frac{m}{m_0 + t - 1}.\tag{3}$$

The degree distribution in this limiting case obeys an exponentially decayed, as $P(k) \sim e^{-k/m}$.

Case B:
$$\phi = t$$
.

This limiting case, with $\phi = t$, means that the local world is the same as the whole network. It is exactly the same as the Barabasi–Albert scale-free model, where the rate of change of the *i*th node's degree is

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2 \cdot t}.\tag{4}$$

The degree distribution in this limiting case follows the power law $P(k) \sim 2m^2/k^3$.

From the above two limiting cases, we can see that if $\phi \approx 0$, the degree distribution is very close to that of Case A above, with $\phi = 0$, as shown in Fig. 6(a). While if $\phi \approx t$, the degree distribution is similar to that of Case B, which follows a power law distribution, as shown in Fig. 6(d). Therefore, if $0 < \phi < t$, the last updating model represents a transition for the degree distribution between the exponential and the power-law distributions, as illustrated by Fig. 6(b) and (c). We find that the



Fig. 6. Degree of the last updating evolving networks with threshold $\phi = 10$; 100; 1000; 10000; $m_0 = 15$, $e_0 = 30$, m = 10, respectively.

degree distribution of the real-life online social network is close to the case, with $\phi = 100$. What is more, the first three distributions are all heavy-tailed, indicating a high heterogeneity between the users.

4.3. Other distributions

The clustering function C(k) is defined as the average of C_i over all vertices with a given degree k. As shown in Fig. 7(a), the trends of the real-life online social network and the last updating evolving network with threshold $\phi = 10$ are nearly the same; their C(k) decays as $\alpha \log(k) + \beta$, with $\alpha < 0$, which is consistent with Ref. [20]. The average nearest-neighbor degree function knn(k) [21,22], which is defined as the average degree of the neighbors of vertices of degree k, also follows a logarithmic distribution, $knn(k) \sim \alpha \cdot \log(k) + \beta$ for these two networks. As shown in Fig. 7(b), knn(k) exhibits a slight downward curvature for both of them, with $\alpha > 0$.

5. Conclusions

The social structure of real-life online social networks reflects both "small-world effect" and skewed degree distributions. However, traditional network evolution models cannot simultaneously depict those two characteristic. In this paper, we will focus on a real-life online social network-Tianya (http://www.tianya.cn), which is one of the largest online social networks in China at present. It is observed that, Tianya community is organized by some continuous active users who can be called forum moderators or super fans. On the other hand, users who have just participated in the discussions on the Internet are more likely to acquire attentions from others. On the basis of these observations, we proposed the concept of "last updating time", which means the time interval from his last interaction on Forum to present.

In our last updating evolution model, the newly coming node connects to *m* nodes, which are selected from its local world with preferential attachment. The local world is selected from existing nodes whose "last updating time" is below the threshold ϕ . In at every time step *t*. The "last updating time" of every node will be updated, depending on their actual interactive behavior. Different network topologies can be acquired by dynamically changing the threshold ϕ .

Computer simulation was implemented in Section 4, in which statistical properties of some artificial networks are analyzed; and we compare the results with the real-life online social network to characterize how they differ or resemble one another. From Fig. 5, we can conclude that if the threshold ϕ is set in a suitable range, the networks we generated have "small world effect". We also indicated that, the last updating model represents a transition for the degree distribution between the exponential and the power-law distributions, in the case of $0 < \phi < t$.



Fig. 7. C(k) and knn(k) of real-life online social network and the last updating evolving network with threshold $\phi = 10, m_0 = 15, e_0 = 30, m = 10.$

Acknowledgments

This work was supported by Jiangsu Innovation Program for Graduate Education (Project NO: CXZZ12_0162) and the Fundamental Research Funds for the Central Universities.

References

- [1] Z. Bu, Z. Xia, J. Wang, A sock puppet detection algorithm on virtual spaces, Knowledge Based Syst. 37 (2013) 366–377.
- [2] Z. Xia, Z. Bu, Community detection based on a semantic network, Knowledge Based Syst. 26 (2012) 30–39.
- [3] http://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg.
- [4] P. Erdos, A. Renyi, On random graphs, I, Publ. Math. Debrecen 6 (1959) 290-297.
- [5] P. Erdos, A. Renyi, On the evolution of random graphs, in: Publication of the Mathematical Institute of the Hungarian Academy of Science, 1960, pp. 17-61.
- [6] D.J. Watts, Small Worlds, in: The Dynamics of Networks Between Order and Randomness, Princeton University Press, Princeton, 1999.
- [7] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
- [8] A.-L. Barabási, R. Albert, H. Jeong, Mean-field theory for scale-free random networks, Physica A 272 (1999) 173–187.
- [9] A.-L. Barabási, R. Albert, H. Jeong, G. Bianconi, Power-law distribution of the world wide web, Science 287 (2000) 2115.
- [10] Ginestra Bianconi, A.-L. Barabási, Bose-Einstein condensation in complex networks, Phys. Rev. Lett. 86 (2001) 5632-5635.
- [11] Xiang Li, Guanrong Chen, A local-world evolving network model, Physica A 328 (2003) 274–286.
- [12] J. Gao, S.V. Buldyrev, S. Havlin, H.E. Stanley, Robustness of a network of networks, Phys. Rev. Lett. 107 (2011) 195701.
- [13] Vicenc Gomez, Andreas Kaltenbrunner, Vicente Lopez, Statistical analysis of the social network and discussion threads in slashdot, in: WWW (2008), Beijing, China, April 21–25, 2008.
- [14] A. Barrat, M. Weigt, On the properties of small-world network models, J. Phys. Eur. B 13 (2000) 547.
- [15] Haibo Hu, Xiaofan Wang, Evolution of a large online social network, Phys. Lett. A 373 (12–13) (2009) 1105–1110.
- [16] Y. Chen, G. Paul, R. Cohen, S. Havlin, S.P. Borgatti, F. Liljeros, H.E. Stanley, Percolation theory and fragmentation measures in social networks, Physica A 378 (2007) 11–19. [Proc. 2006 Budapest Conference on Social Network Analysis].
- [17] Ing-Xiang Chen, Cheng-Zen Yang, Visualization of social networks, in: Handbook of Social Network Technologies and Applications, 2010, pp. 585–610.
 [18] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world'networks, Nature 393 (1998) 440.
- [19] A.-L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509.
- [20] Sara Nadiv Soffer, Alexei Vazquez, Network clustering coefficient without degree-correlation biases, Phys. Rev. E 71 (2005) 051701.
- [21] M. Boguna, R. Pastor-Satorras, A. Vespignani, Absence of epidemic threshold in scale-free networks with degree correlations, Lecture Notes in Phys. 625 (2003) 127.
- [22] N. Matsumura, D. Goldberg, X. Llora, Mining directed social network from message board, in: Proc. 14th WWW'05, ACM Press, New York, USA, 2005, pp. 1092–1093.