A last updating evolution model for online social networks

Zhan Bu [a,b,*], Zhengyou Xia [a], Jiandong Wang [a], Chengcui Zhang [b]

[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

A B S T R A C T

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

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 4 712,859 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