Physica A 432 (2015) 167–179

Contents lists available at ScienceDirect

Physica A

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

A mixing evolution model for bidirectional microblog user networks

Wei-Guo Yuan [a,b,*], Yun Liu [a]

[a] Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China

[b] Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China

HIGHLIGHTS

- We propose a mixing evolution model for bidirectional microblog user networks.
- The community scales follow an exponential distribution.
- Given the node fitness distribution is lognormal.
- Adjusting the parameters of model can generate different simulation networks.

ARTICLE INFO

Article history:
Received 8 April 2014
Received in revised form 2 February 2015
Available online 28 March 2015

Keywords:
Microblog
Complex network
Evolution model
Topology analysis
Community structure

ABSTRACT

Microblogs have been widely used as a new form of online social networking. Based on the user profile data collected from Sina Weibo, we find that the number of microblog user bidirectional friends approximately corresponds with the lognormal distribution. We then build two microblog user networks with real bidirectional relationships, both of which have not only small-world and scale-free but also some special properties, such as double power-law degree distribution, disassortative network, hierarchical and rich-club structure. Moreover, by detecting the community structures of the two real networks, we find both of their community scales follow an exponential distribution. Based on the empirical analysis, we present a novel evolution network model with mixed connection rules, including lognormal fitness preferential and random attachment, nearest neighbor interconnected in the same community, and global random associations in different communities. The simulation results show that our model is consistent with real network in many topology features.

1. Introduction

The substantial interest in the structural and functional properties of complex networks has been partially stimulated by attempts to understand the characteristics of real networks. The small world network model [1] describes a large clustering coefficient and a small average short path length phenomenon in real world. Barabasi and Albert (BA) [2] presented the scale-free network model based on growth and preferential attachment. Based on modularity [3], many community division methods were generated.

* Corresponding author at: Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China.

E-mail address: 10111029@bjtu.edu.cn (W.-G. Yuan).

http://dx.doi.org/10.1016/j.physa.2015.03.043
0378-4371/

Table 1
Two original datasets collected from Sina Weibo.

Dataset	Users	Relationships	Start point
A	118,517	2728,213	Author
B	169,428	2699,411	One celebrity user

More and more extended models of the BA scale-free network have been given [4–9]. A local-world network model [4] was proposed that can adjust the local-world scale parameter to make a degree distribution change between exponential and power-law. Bianconi et al. [6] presented the fitness model, which makes the final selection and connection probability proportional to the product of the node fitness and degree. Considering the normal distribution of fitness coupled with exponential growth in the number of nodes, a stochastic network model [7] was introduced, which can yield a double power-law degree distribution.

Social networks have been extensively studied within the framework of complex networks [10–14]. Recently, with the booming of online social networks (OSNs), many studies have been carried out on network structure and functions [15–21]. The social structure of Facebook friendships network was investigated, including assortativity coefficients and community structures based on different user characteristics [18]. In order to reveal the mechanism of the formation of OSNs, a number of models have been presented. Kumar et al. [22] presented a simple preferential attachment model of OSNs growth, concluding that it is important to consider different classes of users. An updated evolution model [23] was proposed based on the local-world evolving algorithm, which can describe the same characteristics of real-life OSNs characteristics.

Microblogging is an OSN with complex network characteristics, and the most widely used worldwide microblogging site is Twitter. However, most studies have concentrated on the topological characteristics, information diffusion, and user behavior [24–26]. Little attention has been done on the real growth process of microblog user relationship network and reveals the underlying mechanism dominating their evolution. In this paper, we provide one possible model for bidirectional microblog user networks. We find that the number of user bidirectional friends shows lognormal distribution and the community scales of real bidirectional microblog user networks follow an exponential distribution. Furthermore, it is found that the real networks are not only small-world and scale-free but also have some special complex networks properties. Based on these observations, the preferential attachment mechanism does not work well alone. Therefore, our evolution model integrates with multiple connection mechanisms. The simulation results show that this model can fit many topology features of real network. We also provide the distribution of degree analytical results and compare with the power-law form and exponent value by changing with the connection parameters.

The remainder of the paper is organized as follows. Section 2 describes the study dataset and discusses the research motivation. In Section 3, we propose the mixing evolution model based on bidirectional microblog user networks structure. Section 4 presents the simulation results and some discussions. Finally, Section 5 concludes the investigation.

2. Dataset and motivation

2.1. Datasets

Our datasets are obtained from Sina Weibo (www.weibo.com), which is the most popular microblogging site in China. Based on Sina Weibo API, we collected users profiles and relationships by employing a snowball sampling method. This means that we selected one particular user as a starting point and obtained the profile information and relationships, and then crawled the dataset of his friends and followers. We repeat this process until the number of users has grown to the desired size. The profile information includes User ID, the number of friends, followers and bidirectional friends. The relationship among microblog users is a similar format as following User ID with a follower User ID. In order to weaken the impact of the user types, an ordinary user (Network A) and a celebrity user (Network B) were selected separately as the sampling starting points. Finally, two original datasets were obtained as shown in Table 1.

The relationships among microblog users are directed. If user A and user B follow each other at the same time, then they are bidirectional friends. The proportion of bidirectional relationship is about 10% in Sina Weibo. We focus on investigating the bidirectional relationship and ignore the one-way relationships in order to more accurately reflect the social network structure characteristics. Because there are some limitations on data collection from Sina Weibo, the original user relationship data in Table 1 cannot build a complete and closed network. Therefore, the original datasets were further processed to construct two connected networks. First, we find out the user ID intersection between the profile information and relationships, and delete the relationships which are not in the intersection. After removing the isolated nodes, the maximal connected sub-graph can be extracted. At last, we obtain two bidirectional closed networks. To study these networks, we adopt undirected graph in the complex network. Every registered user corresponds to a node and an edge represents a bidirectional relationship between two users. The statistical properties of the two real networks have been listed in Table 2, where V means the number of nodes, E is the number of edges, $\langle k\rangle$ is the average degree, D is the diameter, r is the assortativity coefficient, C is the clustering coefficient, L is the average shortest path length.