

A Practical Simulation Method for Social Networks

Rui Zeng¹, Quan Z. Sheng², Lina Yao², Tianwei Xu¹, Dong Xie³

¹School of Information Science and Technology
Yunnan Normal University, Kunming, China

zengruyn@126.com

²School of Computer Science
The University of Adelaide, Adelaide, SA 5005, Australia

{qsheng, lina}@cs.adelaide.edu.au

³Department of Computer Science and Technology
Hunan University of Humanities, Science and Technology

lgxzy@163.com

Abstract

With the increasing popularity of social networks, it is becoming more and more crucial for the decision makers to analyze and understand the evolution of these networks in order to identify e.g., potential business opportunities. Unfortunately, understanding social networks, which are typically complex and dynamic, is not an easy task. In this paper, we propose an effective and practical approach for simulating social networks. We first develop a social network model that considers the addition and deletion of nodes and edges. We consider the nodes' in-degree, inter-nodes' close degree, which indicates how close the nodes are in the social network, and the limit of the network size in the social network model. We then develop a graph-based stratified random sampling algorithm for generating an initial network. To obtain the snapshots of a social network of the past, current and the future, we further develop a close degree algorithm and a close degree of estimation algorithm. The degree distribution of our model follows a power-law distribution with a "fat-tail". Experimental results using real-life social networks show the effectiveness of our proposed simulation method.

Keywords: Social network, simulation, adjacent matrix, power-law distribution, in-degree, close degree.

1 Introduction

Over the past few years, social networking becomes more and more pervasive. A key aspect of studying social networks is to understand its evolution and the mechanism by which social network structures change over the time. Understanding the dynamic topological map of social networks is important for a variety of different applications such as commercial, security, criminal investigation, etc [9, 23]. Unlike the traditional customer maintenance that always focuses on key customers with higher consumption, in social networks, hot customers are not necessarily high consumption customers. But the loss of these customers may cause the loss of more customers, b-

ecause these customers contact (and influence) many customers. Therefore, using hot customers in a social network to define key customers can compensate the weakness of the approach using high consumption to define key customers. We can easily reveal the hot customers in the topological map of a social network.

The relationships among social network members present huge business opportunities. Based on relationship of customers, a company is able to not only to directly contact its customers, but also to indirectly to do so to other customers. As a result, analysing, understanding, and predicting customers' relationships of social networks attract a significant attention in the research community. However, customer relationships in social networks are highly dynamic. For example, many people who not yet join a social network may do so in the next few weeks; some people may change their relationships; some people may leave the network after period of time etc. These messages are hinted at snapshots of the social network. How to get a snapshot of a social network at a specific time point is the focus of our paper because these snapshots may reveal a lot of business information.

With the continuous expansion of the networks, more and more attention is paid to complex networks. The social network, World Wide Web or biological systems exhibit a structure of a complex network with nodes representing the entities and edges representing some sort of interactions among them. Some recent studies have found that complex networks have the characteristics of scale-free networks. The scale-free networks are graphs with a "fat-tail" degree distribution with considerably higher presence of so called *hubs*, nodes with very high degree, and many of nodes with lower degrees, such that the number of nodes of degree k follows a power-law distribution [1]. Barabasi and Albert first present a scale-free model [2] (also called BA model) that satisfies a large number of real systems. Based on the BA model, in order to provide a more satisfactory description for empirical characteristics of some real networks, many new models have been proposed, including the fitness model [3] and the mixed attachment mechanism's model.

Although these models capture the power-law tail of the degree distribution, it has other properties that may be not accordant with the empirical results of the real networks. A social network is a typical complex network, with a set of people and interactions among the people. Most nodes have very few connections while a small

Copyright © 2012, Australian Computer Society, Inc. This paper appeared at the 1st Australasian Web Conference (AWC 2013), Adelaide, South Australia, January-February 2013. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 144. H. Ashman, Q. Z. Sheng and A. Trotman, Eds. Reproduction for academic, not-for profit purposes permitted provided this text is included.

number of nodes have many connections. Such a network combines addition and deletion of nodes and edges. However, many models focus on growth mechanism and ignore the mechanism of the deletion of nodes. We observe that it is important to consider the minimum size of a network in social network model when we need to delete nodes since unlimited deletion may destroy the network. In order to realize the simulation, we design a social network model which adds nodes and edges at the same preferential attachment probability, but deletes nodes and edges at different anti-preferential attachment probabilities. Preferential attachment is widely used in analyzing the growth of a network [25], and the basic idea of the preferential attachment is that the probability that a new edge involves node x is proportional to the current number of neighbors of x .

We consider two factors of nodes' in-degree and inter-nodes' close degree in preferential attachment probability to establish social network model in order to closely reflect the status of nodes and nodes' relationship in social network with a certain theme. In-degree of nodes denotes the number of edges that connect to nodes. The close degree indicates the closeness between two nodes. In the social network with a certain theme, nodes' status and relationship are determined by nodes' in-degree and inter-nodes' close degree. Because node i that connects to node j should consider not only the in-degree of node j but also the close degree between i and j .

Node's attractiveness is the ability to attract other nodes, which is considered in a few models such as [6]. At time t , a node has same attractiveness to all nodes in these models. It is unreasonable because a node has different attractiveness to different nodes at time t in real social networks. In [17], Liben-Nowell and Kleinberg study the factor of relationship strength in social network. They only consider this factor between existing nodes. In our model, we use the close degree to denote the close extent between nodes instead of attractiveness. Obviously, in a social network, hot nodes may change to common nodes and the relationships between nodes may change over the time as well. In addition, the preferential attachment probability is not only dependent on node's close degree, but also dependent on node's in-degree. In most models, the preferential attachment probability considers the degree of nodes because nodes' connection in these models has no directions. Links connecting nodes in our model have directions, so we consider in-degree of nodes in our model.

Based on our model, we develop algorithms to realize the simulation of social networks. In order to visualize social networks, we first set up social network model which considers the addition and deletion machine of nodes and edges. We then develop an algorithm to generate social network topology maps as snapshots of the social network. Our main contributions are summarized as follows:

- We design a social network model which adds nodes and edges at the same preferential attachment probability, while deletes nodes and edges at different anti-preferential attachment probabilities. Links connecting nodes in our model have directions. It is important for discovering customer re-

lationships. Nodes' connection in most existing models has no directions, such as Riitta Toivonen [9]. For these models, degree is considered in their algorithms. We consider in-degree rather than degree in our model.

- Based on our model, we develop an algorithm of the close degree and an estimation algorithm of the close degree. The estimated value of the close degree is used to generate a snapshot of a social network in the future while the real close degree is used to generate a snapshot of the social network in the past time t or current.
- The initial network of our model is derived from the sample of existing networks. This will make the network simulated by our model close to real situation. In our model, we develop a graph-based stratified random sampling algorithm to generate a sample of existing network.

We validate our approach through the simulation of several real-life social networks including a car fanciers' social network with 25,000 members, a financial customer relationship network with 35,741 members, and a user network in colleges and universities with 25,468 members. We compared the node state of the simulated network with these real social networks. The results show the feasibility of our simulation approach and our approach outperforms the BA model.

The remainder of this paper is organized as follows. In Section 2, we introduce our social network model. In Section 3, we describe the simulation algorithm and discuss the parameter constraints. In Section 4, we report the simulation experiment. Finally, Section 5 overviews the related work and Section 6 gives some concluding remarks.

2 The Social Network Model

In order to simulate social networks, we first develop a model for this purpose. This model includes the new nodes adding mechanism, the old nodes removal mechanism and addition and deletion mechanism of the edges between the old nodes. The four mechanisms reflect the nature of social network development and changes.

A sample of existing networks N_0 has m_0 nodes. We denote by $N(t)$ the size of the network, by $k_i(t)$ the in-degree of node i that denotes the number of edges connected to the node i and by $a_{ig}(t)$ the close degree between node i and node g at time t . Through access to $k_i(t)$ and $a_{ig}(t)$ at time t , we can obtain the corresponding the preferential attachment probability and the anti-preferential attachment probability in order to establish a snapshot of the social network at time t . In each time step, we perform the following four operations:

1. Start with a small number (m_0) of nodes, at every time step, we add a new node with m ($m \leq N(t)$) edges that are connected to m different nodes. The new node g connects to the node i pre-existing in the network with the preferential attachment probability:

$$\prod(k_i) = \frac{k_i(t) + a_{ig}(t)}{\sum_{j=1}^{n-1} (k_j(t) + a_{jg}(t))} \quad (1)$$

The estimated value of $a_{ig}(t)$ is used to generate a snapshot of a social network in the future while the real $a_{ig}(t)$ is used to generate a snapshot of the social network in the past time t or current. The specific algorithm of $a_{ig}(t)$ will be given in Section 3.2.

2. L new edges among old nodes are produced: a node i is selected as an end of a new edge, with the preferential attachment probability given by (1).
3. C old links are deleted: we select a node i as an end of a deleted link with the anti-preferential attachment probability:

$$\prod^*(k_i) = \frac{1}{(N(t)-1)} (1 - \prod(k_i(t))) \quad (2)$$

where $1/(N(t)-1)$ is the normalized coefficient of the probability.

4. E old nodes are deleted: node i is deleted with the anti-preferential attachment probability:

$$\prod^{**}(k_i) = \frac{N(t) - N_{\min}}{N(t)} (1 - \prod(k_i(t))) \quad (3)$$

N_{\min} is the minimum size of the system. The anti-preferential probability is more reasonable for deleting nodes and nodes may be deleted when $N(t) > N_{\min}$.

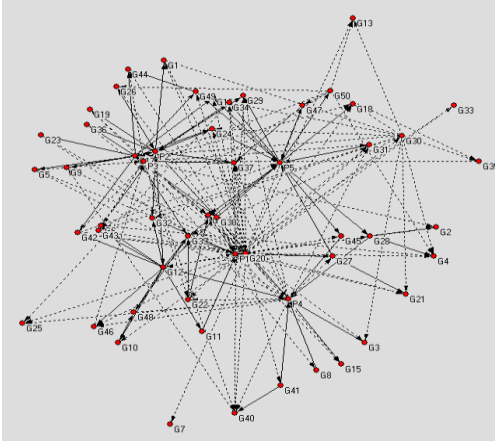


Fig. 1 Topology map of a car fanciers' social network in t1

Fig. 1 and 2 are topology maps of a car fanciers' social network at time t1 and t2. We can discover that the network increases fifty more nodes from t1 to t2. Many edges are deleted such as the edge from G21 to G27 in Fig. 1, while many edges are produced such as G21 to G1 in Fig. 2. Some nodes are deleted such as G56 in Figure 2.

3 The Simulation Algorithm

In this section, we describe the key social network algorithm, as well as discuss the parameters' constrain that could be used for the simulation of social networks.

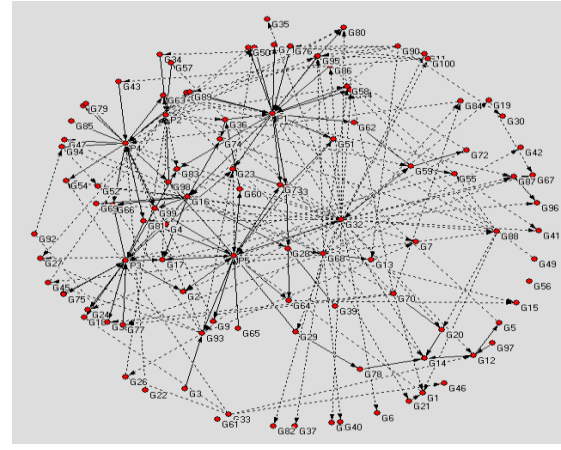


Fig. 2 Topology map of a car fanciers' social network in t2

To model the nodes and the relations of the social network, the *topology map* is defined as an directed graph $G(V, E)$:

Definition 1. $G = (V, E)$ is a network or graph where $V = \{v_1, v_2, \dots, v_n\}$ is the nonempty set of nodes, and E is a set of edges where $E \subseteq V \times V$. E_{ij} is the edge that connects nodes v_i to v_j if $v_i \in V$ and $v_j \in V$.

3.1 Initialization Network

The initial network of our model is derived from a sample of existing networks. This will make the network simulated from our model close to real situations. More recently many experts have studied the map sampling technology of network, such as C. Hubler [18], which proposes an innovative subgraph sampling technique based on the Metropolis algorithm [19] to assess the degree of consistency with graph properties. These methods have focused on the basic properties of map such as relationship strength between nodes. In addition, these methods are very complex and used for different purposes such as graph compression [20, 21] and epidemiology [22]. However, nodes inside a social network are persons or entities that may have many other properties in addition to the links (e.g., hobbies). Therefore, in our work, we design a graph-based stratified random sampling algorithm for generating a sample of existing network. Before introducing our algorithm, we give some definitions as the following:

Definition 2. The sampling fraction, $P(S)$, of a sample S is:

$$P(S) = \frac{F}{N} \quad (4)$$

where N is the number of nodes in the sampling frame and F is the number of nodes in the sample. S is a sample of nodes where $S \subseteq V$.

Definition 3. $G(S)$ is the sample graph of G based on the sample S . That is, $G(S) = (S, ES)$ where $S \subseteq V$ is the vertex set and the edge set $ES = (S \times S) \cap E$.

The historical data of the social network is divided into homogeneous subgroups. There are many techniques to generate subgroups. In our work, we use a clustering algorithm because it is suitable and there are many tools

available. Let S be a random drawn from each subgroup based on $P(S)$. The algorithm is depicted in Fig. 3.

Algorithm 1: Initialization Network
Input: $V=\{v_1, v_2, \dots, v_m\}, F, N$
Output: $a, N(t)$
<ol style="list-style-type: none"> 1. V is divided into V_1, V_2, \dots, V_n based on the clustering algorithm // V_1, \dots, V_n is subgroup of V 2. For $i=1$ to n // n is the number of subgroups of V 3. $P(S)=F/N$ 4. $w=N_{v_i} * P(S)$ // N_{v_i} is the number of nodes of V_i 5. For $j=1$ to w 6. $S \leftarrow v_k$ // v_k is the node that is randomly drawn from V_i 7. End for 8. End for 9. $ES = (S \times S) \cap E$ 10. For $i=1$ to T // T is the number of edges of ES 11. If $E_{kx} \neq \emptyset$ then // E_{kx} is the i edges of ES. 12. $a(k,x)=1$ 13. End if 14. End For 15. $N(t) = C$

Fig. 3 Initialization network algorithm

3.2 Close Degree Algorithm

In our model, we use close degree to denote nodes' close extent. A node has different close degrees to different nodes at time t . It changes over the time and exists in all nodes. Some experts have studied relationship strength between nodes in order to divide network or prediction analysis, such as the work by Liben-Nowell and Kleinberg in [17]. But these methods focus on basic link properties in the map. A social network consists of people and the link between people may imply much more information. In our simulation model, we study this hidden information to obtain close degree between nodes. Based on a theme, through historical data mining, we can obtain the related factors and their weights in the social network. These factors and weights decide the close degree between nodes.

Let $a_{ij}(t)$ be the close degree between node i and node j at time t , f_k the k factor based on a topic's social network at time t and W_k the weight of f_k . The close degree is calculated as the following:

$$a_{ij}(t) = \sum_{k=1}^n (f_k \times W_k) \quad (5)$$

$a_{ij}(t)$ and $k_i(t)$ can be used to generate a snapshot in the past time t or current. In order to simulate future status of the social network, we need to estimate the value of nodes' close degree in the future. We can obtain history volatility of the close degree from historical data. Through the historical volatility of the close degree estimates the future volatility of the social connections. In our approach, we use standard deviation to estimate the future volatility of close degrees. The close degree's estimation can be calculated as follows:

$$A_{ij} = \frac{1}{n-1} \sum_{t=1}^{n-1} (\ln a_{ij}(t+1) - \ln a_{ij}(t)) \quad (6)$$

$$a_{ij}^*(n+1) = a_{ij}(n) \times \left(1 + \sqrt{\frac{\sum_{t=1}^n (\ln a_{ij}(t) - A_{ij})^2}{n-1}}\right) \quad (7)$$

Where A_{ij} is the average of volatility of the close degrees between node i and node j , a_{ij}^* is the estimation of the close degree between node i and node j at the time $n+1$.

3.3 Discussion on Parameter Constraints

A social network is a type scale-free network which in-degree k follows a power-law distribution, where $2 < \gamma \leq 3$ (γ is power index). Through the calculation of the in-degree distribution, we can find out the formula of γ . We can obtain constraints among E, C, L from $2 < \gamma \leq 3$. The in-degree distribution of our model is calculated as follows.

The value of $\sum_j a_{gj}(t)$ varies over the time, so we suppose $\sum_j a_{gj}(t) = At$. The system size $N(t)$ and the total number of in-degree $\sum_j k_j$ vary with the time t as: $N(t) = m_0 + t - E \approx t$, $\sum_j k_j = 2(m + l - c - E)t$ and $\sum_j (k_j + a_{gj}(t)) = (2(m + l - c - E) + A)t$.

We use the mean-field theory [7] to analyze and calculate the in-degree distribution $p(k)$ for various in-degree k . The in-degree k_i of node i is supposed to change continuously. The probability $\Pi(k_i)$ is proportional to the relations at the rate at which k_i changes, so we can have:

$$\begin{aligned} \frac{\partial k_i}{\partial t} &= m\Pi(k_i) + l[\Pi(k_i) + \sum_{j \in \text{linked}(i)} \Pi(k_i)\Pi(k_j)] - \\ &c[\Pi^*(k_i) + \sum_{j \in \text{linked}(i)} \Pi^*(k_i) * \Pi^*(k_j)] - E\Pi^{**}(k_i) \\ &\approx m\Pi(k_i) + l[2\Pi(k_i) + (\Pi(k_i))^2] - c \left[\frac{2(1-\Pi(k_i))}{t} - \right. \\ &\left. \frac{(1-\Pi(k_i))^2}{t^2} \right] - E \frac{N(t)-N_{\min}}{N(t)} (1 - \Pi(k_i)) \\ &\approx (m + 2l)\Pi(k_i) - \frac{2c}{t} - E \frac{t-N_{\min}}{t} \\ &= \frac{(m+2l)(k_i+a_i(t))}{[2(m+l-c-E)+A]t} - \frac{2c+EN_{\min}}{t} - E \\ &\approx \frac{(m+2l)(k_i+a_i(t))}{[2(m+l-c-E)+A]t} - \frac{2c+EN_{\min}}{t} \\ &= \frac{(m+2l)k_i}{[2(m+l-c-E)+A]t} + \frac{(m+2l)a_i(t) - (2c+EN_{\min})[2(m+l-c-E)+A]}{[2(m+l-c-E)+A]t} \end{aligned}$$

Suppose

$$\beta = \frac{(m+2l)}{2(m+l-c-E)+A}$$

$$B = \frac{(m+2l)a_i(t) - (2c+EN_{\min})[2(m+l-c-E)+A]}{[2(m+l-c-E)+A]t}$$

We have

$$\frac{\partial k_i}{\partial t} = \beta \frac{k_i}{t} + B \frac{1}{t} \quad (8)$$

With the initial condition that node i is added to the system at time t with the in-degree $k_i(t_i) = m$.

$$k_i(t) = \left(m + \frac{B}{\beta}\right) \left(\frac{t}{t_i}\right)^\beta - \frac{B}{\beta} \quad (9)$$

Every step, we add one node into the network, the probability density of t_i is

$$P(t_i) = \frac{1}{m_0 + t}$$

We can obtain

$$\begin{aligned}
 P(k_i(t) < k) &= P(t_i > \left(\frac{m+\frac{B}{\beta}}{k+\frac{B}{\beta}}\right)^{\frac{1}{\beta}} t) \\
 &= 1 - \frac{t}{m_0+t} \left(\frac{m+\frac{B}{\beta}}{k+\frac{B}{\beta}}\right)^{\frac{1}{\beta}}
 \end{aligned}$$

Where

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k}$$

So we can obtain

$$P(k) = \frac{t}{m_0+t} \frac{1}{\beta} \left(m + \frac{B}{\beta}\right)^{\frac{1}{\beta}} \left(k + \frac{B}{\beta}\right)^{-\left(1+\frac{1}{\beta}\right)} \quad (10)$$

For $m_0 \ll t$, So

$$P(k) \approx \frac{1}{\beta} \left(m + \frac{B}{\beta}\right)^{\frac{1}{\beta}} \left(k + \frac{B}{\beta}\right)^{-\left(1+\frac{1}{\beta}\right)}$$

Where

$$\gamma = 1 + \frac{1}{\beta} = \frac{2(m+l-c-E)+A}{(m+2l)} \quad (11)$$

$$2 < \gamma = \frac{3m+4l-2c-2E+A}{(m+2l)} \leq 3 \quad (12)$$

Values of A can be obtained through Equation 5. If E, C, and L satisfy Equation 12, this model could self-organize into a stable network. In our algorithm, E, C and L are generated in a random way, but they must satisfy Equation 12.

4 Experiments

To validate the efficiency and effectiveness of the proposed social model and our simulation method, we implemented the algorithm using C++, Matlab and PAJEK. We extracted the three sets of real historical data of the social network test the simulation accuracy of our model. These data are provided by a telecommunication company. These social networks consist of nodes that are the phone number and edges that are telephone contacts between two nodes. The three social networks are (1) the car fans network with 25,000 nodes, (2) the financial customer relationship network (referred to as financial network) with 35,741 nodes, and (3) user network in colleges and universities (referred to as university network) with 25,468 nodes.

Each set of data constitutes a social network G (see Definition 1). We use the data of G to obtain the related factors and their weights of close degree (see Section 3.2) using the factor analysis method and homogeneous subgroups of social network (see Section 3.1) using cluster analysis. In our experiment, there are 500 nodes in each N_0 ($N_0 = \langle V_0, E_0 \rangle$, where $V_0 \neq \emptyset$ and $V_0 \subseteq V$, and E_0 is a set of edges where $E_0 \subseteq V_0 \times V_0$ and $E_0 \subseteq E$). We use stratified random sampling methods to extract 1,500 real historical data that belong to V but not to V_0 as the input data set V_I to simulate the prediction social network G^* ($G^* = \langle V^*, E^* \rangle$, where $V^* = V_I \cup V_0$, $E^* \subseteq V^* \times V^*$) that grow following the four steps of our model (see Section 2).

4.1 Simulation Degree Description

Researchers usually learn a network topology according to the degree distribution. Therefore, we will look into the

degree distribution of simulation network to estimate whether the simulated social network reflects the real structure of the network.

The map of degree distribution of the car fanciers' social network is generated by a program written in Matlab. The in-degree map and out-degree map are depicted in Fig. 4 and Fig. 5, respectively. These maps show that the degree distribution of the simulated network follows a power law distribution, where $\gamma = 3.0$ in Fig. 4, $\gamma = 2.8$ in Fig. 5. We can see from the figures that the number of nodes whose degree is less than 4 is very large, whereas the number of high degree nodes is very small.

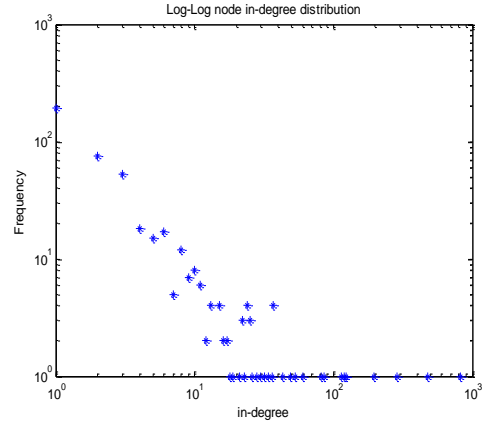


Fig. 4 The in-degree distribution map

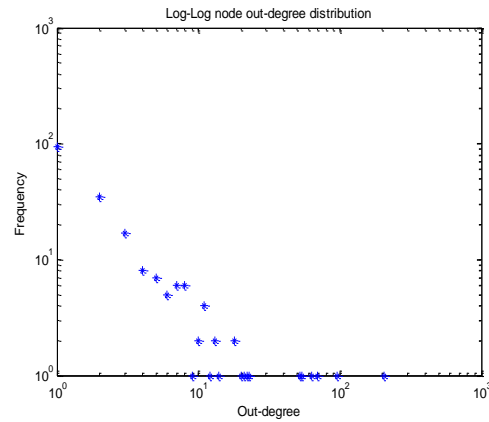


Fig. 5 The out-degree distribution map

In order to validate our simulation, we produced a degree statistical chart of the real historical data of the car fancier's social network, shown in Fig. 6. From the figure we can see that the number of nodes whose in-degree is less than 4 is about 53.2%, so the majority nodes have very small connections. The degree of 1.8% nodes nearly reaches 40. This complies with our simulation result. The in-degree and out-degree distribution of the simulated social network follows a power-law distribution with a "fat-tail". The simulated social network model could self-organize into a scale-free network when E, C and L satisfy the equation (12). We can see that our proposed social network model reflects well the structure of real social networks.

4.2 Simulation Topology Map

A program written in C++ based on our model produces an input file that includes vertices, relations and time points. Social network topology map is drawn when the input file is read into PAJEK. In our simulation experiment, we exploited a sample of real historical data of the social network collected as N_0 to generate the simulated social network. At every time interval, a number of real historical data from the input data set were pushed into the simulated algorithm to generate simulation social network.

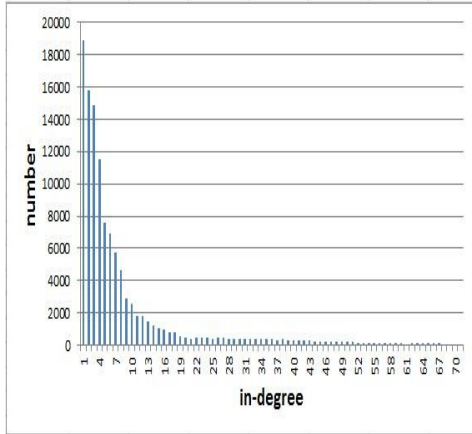


Fig. 6 The in-degree statistical chat

We used the close degree algorithm in our model to generate the topology of the social network in the past time t or current. In order to show the effectiveness of our simulation model, we randomly captured two topography maps of consecutive time interval in Figure 7 and Figure 8. Figure 7 shows the topology map of the simulated car

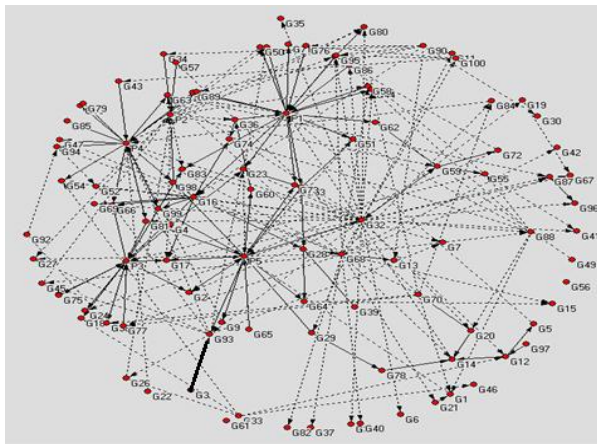


Fig. 7 The topology of car fans network in t_1

fanciers' social network at time t_1 . Figure 8 shows the topology map of the simulated car fanciers' social network at time t_2 (t_2 after t_1). From these topology maps we can intuitively reveal the status of every node and relationship in certain time points such as node G158 and G34 (circled ones in the bottom), which are deleted (to make it easy for readers to compare, this node is still kept in the figure), node G122 is hot node in Fig. 8 and G3 connected to G93 in Fig. 7 (see the thick black line). From topology map of consecutive time interval, we could discover the relationship changes of nodes, such as G3 connected to G93 at time t_1 (see Fig. 7), while it connects to G141 at t_2 (see Fig. 8).

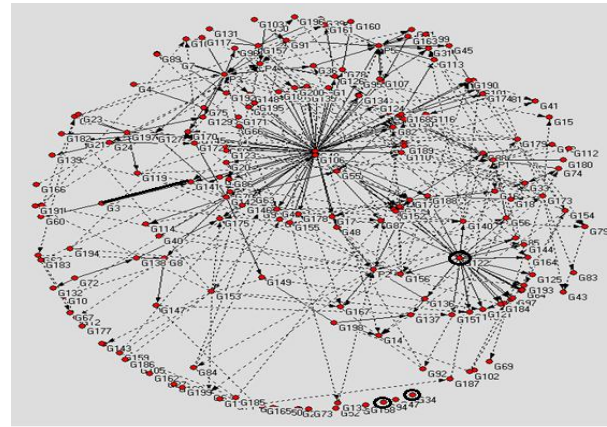


Fig. 8 The car fanciers' network in t_2

To get a snapshot of the car fanciers' social network in the future, we employed the proposed estimation of close degree algorithm. The result is shown in Fig. 9. From Fig. 9, we can discover the status of each node and their relationships in the future. For example, G57 will be deleted, G99 will connect to G170 and G131 will become a hot node (see Fig. 9). This is very important for practical analysis and prediction in many applications.

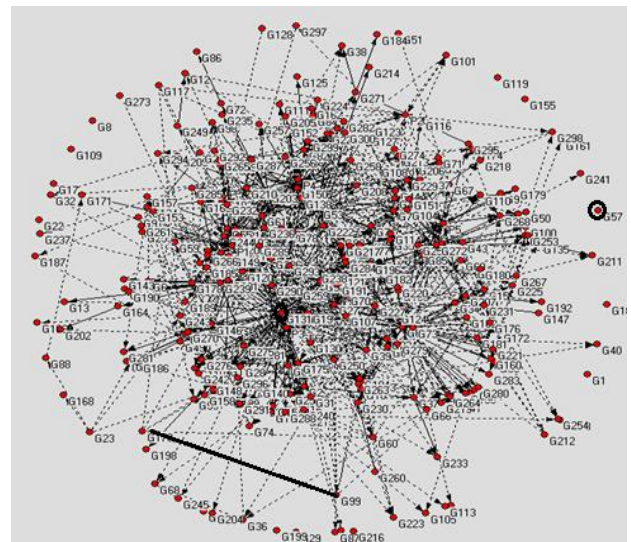


Fig. 9 The estimated car fanciers' network at a future time-point

In order to measure the prediction accuracy, we define the accuracy rate of the prediction as the following

$$Ar = \frac{S_c}{S_I} \quad (13)$$

where S_I represents the number of nodes in the input data set, S_c denotes the number of the accuracy connected nodes that is belong to the input data set. If $v_j \leftarrow v_i \in E^*$ is same as $v_j \leftarrow v_i \in E$, $v_i \in V_I$, $v_j \in V^*$, $V_I \cap V_0 = \emptyset$, $V_I \cup V_0 = V^*$, v_i is properly connected node.

In order to confirm our approach is valid and has better accuracy than that of the BA model, we not only generate the car fanciers' network, the financial network and the university network based on our approach, but also generate the financial network based on the BA model (referred to as the network of the BA model). Simulation accuracy rate is shown in Table 1. The results show that

three networks simulation of current or past time average accurate rate reaches 72.37%, the simulation of future average accurate rate reaches 64.53%. The accuracy rate of the network of the BA model is only capped at 44.84%. Because BA model does not have a future state prediction, the simulation accuracy rate of current or the past and the future period of the network of the BA model are the same. The accuracy rate of our approach is higher than that of the BA model. The accuracy rates of the three networks are very close. As a result, the accuracy rate of our approach is stable. Because of the noise data, it is difficult to precisely calculate the close degree and estimated value of the close degrees. With more accurate and complete historical data, our approach will produce better result in terms of simulation accuracy.

Social networks	Car fancier's network	Financial network	University network	The network of the BA model
Accurate rate of the past	72.3%	69.2%	75.6%	44.84%
Accurate rate of the future	67.6%	60.8%	65.2%	44.84%

Table 1 simulation accuracy rate of social networks

5 Related Work

Despite the increasing popularity of social networking, there are very few tools available to support the simulation of social networks. D'Angelo and Ferretti develop a scale-free net simulator called PaScas [4]. The model can add local characteristics to nodes and implement network topology, which can simulate heterogeneous network environments. Ren and Li [5] propose an algorithm, named RX, which can simulate scale-free networks. However, this model does not consider attachment possibility [1]. There are some simulations in gene fields, such as Banjo [6], ARACNE [7] and NETI [8] etc. Kincaid and Alexandrov show how to build a scale-free network to simulate air transportation networks. All of these simulation methods focus on the algorithm of the BA model [2]. In the BA model, there is no nodes deletion mechanism. In our simulation, we consider addition and deletion of nodes and edges mechanism that contain close degree factors and in-degree factors. In these existing simulation approaches, the results of simulation can be only used for analyzing structural and properties of networks.

Dorogovtsev and Mendes [2] develop a model which interleaves the addition of nodes and edges with a uniform preferential attachment probability. Chung and Lu [3] and Cooper et al. [4], independently, investigate a dynamic model in which nodes and edges were deleted by uniform preferential attachment probability. Narsingh Deo [24] propose a dynamic random graph model which interleaves addition of nodes and edges with a preferential deletion of nodes, but there was not limit of size of the network when nodes were deleted. We observe that it is important to consider the minimum size of a network in social network model when deleting nodes since unlimited deletion may destroy the network. In order to realize the simulation, we design a social network model which adds nodes and edges at the same preferential attachment probability, but deletes nodes and edges at difference anti-preferential attachment probabilities.

6 Conclusion

With the rapid adoption of the social networking, it becomes more and more important for decision makers to analyze and understand social networks. Simulation-based methods are widely regarded as a sufficient way for understanding these complex, dynamic systems. In this paper, we propose a social network simulation model that takes into account the addition and deletion of nodes and edges. The nodes' in-degree, inter-nodes' close degree and limit of the network size are considered in the social network model. In order to visualize social networks, we develop i) a graph-based stratified random sampling algorithm to generate initial network, and ii) a close degree algorithm and close degree of estimation algorithm to get snapshots of the social network in the past, current and the future. Experiments show that the in-degree distribution map and the out-degree distribution map of our model follow the power-law distribution with a "fat-tail". Experiments also confirm that our approach can effectively simulate social networks.

In the future, we plan to do more experiments to further validate the performance of our approach. Optimization of simulation performance is a main focus of our future research work. We will also consider the event-driven changes of social networks. Such changes do not necessarily cause adding or deletion of nodes, but can lead to the structural evolution in the social networks.

7 Reference

- [1] S.N. Dorogovtsev, J.F.F. Mende. (2000): Scaling behaviour of developing and decaying networks. *Europhys Letters* **52** (1):33–39.
- [2] A.-L. Barabasi and R. Albert. (1999): Emergence of scaling in random networks. *Science* **286** (15):509.
- [3] Bianconi G and Barabási A L. (2001): Competition and multiscaling in evolving networks. *Phys* **54**(4): 436-440.
- [4] Gabriele D'Angelo and Stefano Ferretti. (2009): Simulation of Scale-Free Networks. *ICST21*, Belgium.
- [5] Wei Ren and Jianping Li. (2009): A fast algorithm from simulating scale-free networks. *ICCT*.
- [6] Yu J, Smith V A, Wang P P. (2004): Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* **20**(18):3594-3603.
- [7] Basso K, Margolin A A, Stolovitzky G. (2005): Reverse engineering of regulatory networks in human B cells. *Nature Genetics* **37**(4):382-90.
- [8] Novikov E and Barillot E. (2008): Regulatory network reconstruction using an integral additive model with flexible kernel functions. *Bmc Systems Biology*.
- [9] Riitta Toivonen, Jukka-Pekka Onnela. (2006): A model for social networks, *Physica A* 851–860.
- [10] A.-L. Barabási, R. Albert, H. Jeong. (1999): Mean-field theory for scale free random networks, *Physica A* **272**:173–187.
- [11] M.E.J. Newman. (2003): The structure and function of complex networks, *SIAM Rev* **45**:167–256

- [12] Hongan Che, Jifa Gu. (2004): Scale-free network and system scientific significance. *The System Engineering Theory and Practice* **4**:11-16.
- [13] Sarshar N, Roychowdhury V. (2004): Scale-free and stable structures in complex ad hoc networks. *Physical Review E* **69** (2):357-368.
- [14] Dorogovtsev SN, Mendes JFF and Samukhin AN. (2000): Structure of growing networks with preferential linking. *Physical Review Letters* **85**:4633-4636.
- [15] Cooper C, Frieze A, Vera J. (2004): Random deletion in a scale-free random graph process. *Internet Mathematics* **1**(4):463-483.
- [16] R. Albert, A.-L. Barabási. (2002): Statistical mechanics of complex networks. *Phys* **74**:47-97.
- [17] David Liben-Nowell, Jon Kleinberg. (2007): The Link-Prediction Problem for Social Networks. *The American society for information science and technology* **58**(7):1019-1031.
- [18] C. Hubler, H.-P. Kriegel, K. Borgwardt, and Z. Ghahramani. (2008): Metropolis algorithms for representative subgraph sampling. *ICDM'08* 283-292.
- [19] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. (1953): Equation of state calculations by fast computing machines. *Phys* **21**(6):1087-92.
- [20] T. Feder and R. Motwani. (1991): Clique partitions, graph compression and speeding-up algorithms. *J. Comp. and Sys. Science* 123-133.
- [21] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai. (2000): The large-scale organization of metabolic networks. *Nature* **407**(6804):651-654.
- [22] S. Goel, Matthew, and J. Salganik. (2009): Respondent-driven sampling as markov chain monte carlo. *Medicine*.
- [23] A. Squicciarini, S. D. Rajasekaran, and M. C. Mont. (2011): Using Modeling and Simulation to Evaluate Enterprises' Risk Exposure to Social Networks. *IEEE Computer* **44**(1): 66-73.
- [24] Narsingh Deo, Aurel Cami. (2007): Preferential deletion in dynamic models of web-like networks. *Information Processing Letters* **102**:156-162.
- [25] Michael Mitzenmacher. (2001): A brief history of lognormal and power law distributions. *The Allerton Conference on Communication, Control, and Computing* 182-191.
- [26] Peter Pollner, Gergely Palla, and Tamas Vicsek. (2006): Preferential attachment of communities: the same principle, but a higher level. *Europhysics Letters* **73**(3):478-484.