

An Interest-Oriented Network Evolution Mechanism for Online Communities

Caihong Sun and Xiaoping Yang

School of Information, Renmin University of China, Beijing 100872, P.R. China
chsun.yang@ruc.edu.cn

Abstract. Online communities involve lots of interactions among members. Those interactions are usually shown as “friend links” between community members, and an online community could be seen as a network of members. The structures and properties of community network formed are important to the information diffusion in communities. Hence, the network formation mechanisms attract great interest in recent years. In this paper, an interest-oriented network evolution mechanism is proposed to study the growth and evolution of social network of online communities. Agents in an online community tend to choose those who share similar interests with them to be their friends. In our model, we define n different interest categories; they could refer to sports, movies, music, literatures and so on. Each agent has an interest vector (V_1, V_2, \dots, V_n) to indicate its interests.

Keywords: *Network evolution mechanism, Online communities, Social network, Information diffusion, Friendship network*

1. INTRODUCTION

In recent years, online communities (i.e. Wikis, Blogs, Forums, etc) have attracted great interest and become the important knowledge sharing resources. Online community or virtual community is a group of people communicating or interacting with each other by means of information technologies, typically the Internet, rather than in person. In brief, an online community is a kind of computer-supported social network (CSSN) [1, 2], in which members and their “friend links” form social networks. How to form the structure of an online community from the beginning? What are the mechanisms of network evolution? Many researchers have put their efforts on studying these network evolution mechanisms ever since the random graph model of Erdos and Renyi could not explain some features of network structure. Degree distribution of random graphs follows a Poisson law, whereas many empirical social networks exhibit power law degree distributions. Watts and Strogatz [3], Barabasi and Albert [4] proposed small world and scale-free network evolution models respectively. They provided new emerging simulation techniques to investigate the dynamics of a social network on the evolution of the network itself overtime.

In 1995, Holland first brought forward the concept of “tag”. A tag works like a flag that identifies one group of users from another. Holland assumed that arbitrary, evolving tags could facilitate selective interactions and thereby be helpful for aggregation and boundary formation [5]. In Riolo’s work published on Nature, they used computer simulation methods and demonstrated that tag-based mechanism could lead to the emergence of cooperation even when the agents do not receive reciprocity and are unable to observe or remember others’ actions [6].

In this paper, we proposed an interest-oriented network evolution mechanism based on the idea of tag and Schelling Segregation Model [7]. The model is an agent-based model, in which community members are simulated as agents who can decide whether to link other agents as its friends or not, according to their interest similarities. The main idea of this model is that agents in an online community tend to choose those who share similar interests with them or have high reputation value to be their friends. Agents’ interests could be affected by their friends’ interests by being conform to one another. This simulates the trends that agents learn from their friends. In the proposed model, n different interest categories are defined; they could refer to sports, movies, music, literatures etc. Each agent has an interest vector (v_1, v_2, \dots, v_n) to indicate its interests. If an agent is interested in interest category i , then $v_i = 1$, otherwise $v_i = 0$. The more two agents share the same interest categories, the more the two agents’ interests are similar.

The rest of this paper is organized as follows: In Section 2, an interest-oriented network evolution mechanism is presented to simulate the evolution of the social network structures of online communities. Section 3 discusses the experimental design and demonstrates some simulation results. Section 4 concludes this paper and discusses future work.

2. AN INTEREST-ORIENTED NETWORK EVOLUTION MECHANISM

The interest-oriented network evolution mechanism simulates the phenomena in real world that people on internet tend to choose that who share the similar interests with themselves to be their friends. But usually people have many interests, how to simulate these interests and define the interest similarity? We define n different interest categories; they could refer to sports, movies, music, literatures etc. Each agent has an interest vector (v_1, v_2, \dots, v_n) to indicate its interests. If an agent is interested in interest category i , then $v_i = 1$, otherwise $v_i = 0$. In addition, an agent could specify its interest weight for each interest element as (w_1, w_2, \dots, w_n) . The similarity between agent i and j is defined as Equation (1).

$$IS(i, j) = 1 - \frac{\sum_{h=1}^n w_{ih} |v_{ih} - v_{jh}|}{\sum_{h=1}^n w_{ih}} \tag{1}$$

Where $(w_{i1}, w_{i2}, \dots, w_{in})$ is the interest weight vector of agent i . $0 \leq w_{ih} \leq 1$, for any agent i , and $h=1, \dots, n$. Agent i and j may have different interest weight vector, i.e. $(w_{i1}, w_{i2}, \dots, w_{in}) \neq (w_{j1}, w_{j2}, \dots, w_{jn})$.

Moreover, each agent has a threshold (denoted as T) which is the degree of similarity by which an agent chooses its friends. Agents have their own private similarity tolerance degree. Our mechanism is that an agent randomly selects some agents as its friend candidates, but only those who have similarity greater than its similarity tolerance threshold will become its friends.

The proposed interest-oriented network evolution mechanism follows three rules:

- Rule 1. Growth of agents: Starting with a small random network (with m_0 nodes, and probability p), and at each time step, we add m new agents. The interest vector, weight vector and threshold of the new agent are randomly generated. All agents follow Rule 2 to add their links.
- Rule 2. Growth of links: At every time step, each new agent selects its own “friend” for g ($g \geq 1$) times. The selection includes two steps: first, the agent A randomly gets another agent in the network; second, it compares the selected agent’s interests and its own interests; if the interest similarity is greater than its threshold (T_A), then one link to the selected agent is added, otherwise the agent will not add the link.
- Rule 3. Learning from friend’s interest: At every time step, an agent who adds a new link will adapt its own interest vector according to the interest vector of its new friend. The agent will add a new interest element which it has no such interest but its friend has according to its interest weights.

3. EXPERIMENT DESIGN

In our experiments, the number of interests (denotes as k) of each agent follows normal distribution, i.e., $k = N(\mu, 1)$, where $\mu < n$. For an agent with k interests, we randomly select k interests to form its interest vector. The more two agents share the same interest categories, the more the two agents’ interests are similar. The interest weight vector is used to provide an interface to specify the importance of its interest for a member in an online community. To simplify the simulations, let all the interest weights equal 1, in other words, every interest category is treated equally by all the agents.

The similarity threshold T follows the uniform distribution with the range of (0.1, 0.5) if n equals 10. Every agent has its own similarity threshold. If n equals 1, then the interest-oriented mechanism is degraded into the tag-based mechanism proposed in [8], if we do not treat the interest element as a binary but a real number in $[0,1]$.

In order to study the effectiveness of interest-oriented network evolution mechanisms, we plan to do the following experiments.

3.1 Experiment 1

What are the degree distribution, average path length and clustering coefficient of the networks generated by the proposed mechanism?

Parameters in this experiment are:

The number of agents: $N=10000$

The number of interest categories: $n=2, 5, 10, 15$

The mean μ : $\mu=1, 2, 5, 8$ Note: the number of interests (denotes as k) of each agent follows normal distribution: $k = N(\mu, 1)$.

The initial random network: $m_0=10, p=0.2$

The number of agents who can gain links: $m=1, 2, 5, 10$

The selection chance: $g=1, 5, 10, 100, 1000, N$

Threshold follows the uniform distribution with the range of $[0, 0.5]$, $[0.1, 0.5]$, etc.

In this experiment, we study the degree distribution of network by changing n , k , m , and g respectively. The parameter g denotes the friend search area, parameter m represent the scale of the new agents in each time step.

Degree distribution, average path length and clustering coefficient are three important properties researched in social network: Degree distribution describes the probability distribution of degrees in a network; average path length can describe how fast information can travel in a network; and clustering coefficient of a network is used to describe how closely friends are clustered in a network.

3.2 Experiment 2

What will happen if Rule 3 (learning from friends' tags) is removed from our mechanism?

Based on parameters set up in the experiment 1, by deleting Rule 3 in the proposed mechanism to examine how the degree distribution, average path length and clustering coefficient change.

3.3 Experiment 3

What are the difference on degree distribution, clustering coefficient and average path length between empirical data and networks generated from the interest-oriented mechanism?

The main aim of experiment 3 is to compare the network structure generated from the interest-oriented mechanism and some empirical networks existing in real world.

4. CONCLUSIONS AND FUTURE WORK

In this paper, an interest-oriented network evolution mechanism is presented. The main aim of this mechanism is to generate a higher-quality community in which friend links are more evenly distributed, distances between members are smaller and members are more closely clustered by interests. Three experiment designs are given to study the effectiveness of the proposed mechanism. Comparing with the tag-based mechanism and other network evolution mechanisms, such as preferential attachment [4,9,10], small world model, and the interest-oriented network could have multidimensional concerns when adding a new link.

Our future work includes: (1) analyzing the experimental results based on the experiment designs proposed in section 3; (2) studying the emergent properties of the mechanism which combines interest-oriented mechanism with preferential attachment mechanism; (3) examining the interest weight effects; and (4) considering the life cycle of the agents. We will examine the effects on friend networks by breaking some friend links and introducing agents' life span.

REFERENCES

1. B.Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia and C. Haythornthwaite, Computer Network as Social Networks: Collaborative Work, Telework, and Virtual Community, *Annual Review of Sociology*. Volume 22, pp.213-238, (1996).
2. L. Garton, C. Haythornthwaite and B. Wellman, Studying Online Social Networks, *Journal of Computer Mediated Communication*. Volume 3, Number 1, (1997).
3. D.J. Watts and S.H. Strogatz, Collective dynamics of 'small-world' networks, *Nature*. Volume 393, pp.440-442, (1998).
4. A.L. Barabaasi and R. Albert, Emergence of scaling in random networks, *Science*. Volume 286, pp.509-512, (1999).
5. J. Holland, *Hidden Order: How Adaptation Builds Complexity* (Addison Wesley, 1995).
6. R. Riolo, R. Axelrod, and M.D. Cohen, Evolution of cooperation without reciprocity, *Nature*. Volume 414, pp.441-443, (2001).
7. T.C. Schelling, *Micromotives and Macrobehavior* (W. W. Norton and Co, 1978), pp.147-155.
8. C. Sun, Y. Xu and X. Yang, A tag-based network evolution mechanism for online communities, *the third international conference on natural computation* (Forthcoming, 2007).
9. C. Roth, Generalized preferential attachment: towards realistic social network models, in *Proceedings of Workshop on Semantic Network Analysis at the ISWC* (2005).
10. M. Dell'Amico, Highly-clustered networks with preferential attachment to close nodes, *European Conference on Complex Systems* (Oxford, 2006).