Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Modeling and Simulation on Information Propagation on Instant Messaging Network Based on Two-layer Scale-free Networks with Tunable Clustering

Yu WU [1] Yanrong YANG [2] Huanzheng WU [3]
Institute of Web Intelligence
Chongqing University of Posts and Telecommunications
Chongqing, China, 400065
1 wuyu@cqupt.edu.cn; 2 yangyanrongyyr@163.com;
3 wuhuanzheng1@126.com

Gongxiao WANG [4]
Department of Logistical Information Engineering
Logistical Engineering University of PLA
Chongqing, China, 400065
4 wgx_ss@163.com

***Abstract*—Since Instant Messaging Network is one of the most important ways to propagate information, in order to monitor and forecast the propagation behaviors on Instant Messaging Network, it is necessary to study information propagation nature, discipline and methods. Scale-free network with tunable clustering can be used to build instant messaging. Based on it, we propose a new two-layer model of Instant Messaging Network, and formulate its information propagation rules as well. We have simulated this model and observed the diversification of group numbers, clustering coefficient, spreader and stifler. Results show that the more the number of group network is, the larger cluster coefficient and the higher maximum spreader and the higher final value of stifler are. Thus group network and cluster coefficient effect the information propagation.**

***Keywords*—Instant Messaging, Information propagation, Clustering, Scale-free network**

I. INTRODUCTION

In the information age, information propagation plays an increasingly prominent role, which impacts on international community's politics, economy, technology, culture and etc. Instant Messaging Network has been one of the important ways to propagate information between mass, and is the main means to communicate in corporation intranet as well. There are many well-known Instant Messaging tools recently, such as MSN, Yahoo!, Skype, Tecent QQ, Fetion, Neteast popo, and so on. According to the report of *the Second Quarter of 2008 Chinese Instant Messaging Market Monitoring,* the active accounts number of Instant Messaging Market in China has reached 426 million in the second quarter of 2008. Especially, the active accounts number of Tecent QQ has reached 341 million, which shows that Tecent QQ has been the main platform to communicate instantly in China. In order to monitor and forecast the Instant Messaging Network, researches on information propagation nature, discipline and methods have practical significance.

Watts and Strogatz[1] proposed small-world network model (SW for short), while Barabasi and Albert[2][3] proposed scale-free network model (BA for short), and both are basic models to research complex network. Changyu Liu [4] studied public opinion propagation in relationship network based on small-world network model. They built a basic public opinion propagation simulation model. In addition, Moreno[5] built rumor spreading model based on scale-free network.

By Obtaining the statistic data of the nioki.com instant messaging network, Smith[6] found that instant messaging network exhibited all the characteristics of a scale-free network. Moreover, it had the evident characteristic of a small-world network as well. The degree distribution of small-world network with high clustering coefficient didn't obey power-law distribution, while the degree distribution of BA scale-free network with very low clustering coefficient didn't obey power-law distribution. Holme and Kim[7] introduced scale-free network model with tunable clustering with power-law distribution and higher clustering coefficient. So, the scale-free network with tunable clustering coefficient model can be used as a model to study the instant messaging network.

Recently, the above model is mainly used to build the relationship network in real life, namely, the one-layer model. However, in instant messaging network, besides building the relationship network between one user and the other users in the buddy list, it is necessary to build the relationship network between the user and the group with the same aims or interests called group network. Therefore, we propose a new two-layer network model. This two-layer model is applicable to study the information propagation on the network with group (for instance, Tecent QQ).

The paper includes four main sections. Section II introduces some related work including scale-free network with tunable clustering and clustering coefficient. Section III presents instant messaging network based on two-layer model and its algorithm. Section IV describes and performs one-layer model and two-layer model simulations and does the experiments when the users connect with group networks and the users break the link(s) with group network(s), then analyzes the results of the simulations. Section V provides concluding remarks.

II. RELATED WORK

A. Scale-free Network with Tunable Clustering

Scale-free network with tunable clustering solves the deficiencies of small-world network model and scale-free

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

network model respectively. Based on BA scale-free network model, triad formation is added to build scale-free network with tunable clustering.

First of all, network is defined as an undirected graph G=<V, E>, where V is the set of vertices and E is the set of edges. An edge connects pairs of vertices in V and not more than one edge may connect a specific pair of vertices. The basic algorithm are as follows [6][7].

Step 1 Initial network: To start with, the network consists of m_0 vertices and has no edges.

Step 2 Growth: One vertex v with m edges is added at every time step. Time t is defined as the number of time steps.

Step 3 Preferential attachment (PA): Each edge of a new node v is then attached to an existing vertex w with the probability P_w to its degree:

$$P_w = \frac{k_w}{\sum_{v \in V} k_v}$$

where k_w is the degree of vertex w and k_v is the degree of vertex v.

Step 4 Triad formation (TF): If an edge between v and w was added in the previous PA step, then add one more edge from v to a randomly chosen neighbor of w. If there remains no pair to connect, i.e., if all neighbors of w were already connected to v, then do a PA step instead.

When a vertex v with m edges is added to the existing network, one PA step is performed firstly, then perform a TF step with the probability P_t or a PA step with the probability 1-P_t. The average number m_t of the TF trials per added vertex is then given by $m_t = (m-1)P_t$.

B. Clustering Coefficient

Clustering coefficient reflects the aggregation of the nodes, which is one of the statistical properties to the complex network theory. Therefore, if a vertex v_i has k_i neighbors, $k_i(k_i-1)/2$ edges could exist among the nodes within the neighborhoods, E_i stands for the total number of edges between v_i and its neighborhoods. Thus, the clustering coefficient can be defined as:

$$C_i = 2E_i/(k_i(k_i-1)) \quad (1)$$

The clustering coefficient for the whole network is given by Watts and Strogatz[1] as the average of the clustering coefficient for each vertex:

$$C = \frac{1}{n}\sum_{i=1}^{n} C_i \quad (2)$$

Obviously, in the scale-free network with tunable clustering, as the greater P_t reaches, the higher clustering coefficient is. When P_t = 0.5, clustering coefficient will be about 0.28

III. INSTANT MESSAGING INFROMATION PROPAGATION MODEL

A. Work Idea

Firstly, instant messaging network is considered as undirected graph in the graph theory view, in which the user is regarded as nodes. Then, two-layer model of instant messaging network is built using scale-free network with tunable clustering. Secondly, initial status is allocated to each node, information propagation rules are formulated as well. Finally, simulation is performed and analyzed the results. The flow chart is shown in Fig.1.

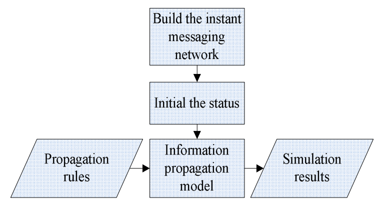

Figure 1 Instant massaging network information propagation flow chart

B. Our Two-layer Instant Messaging Network

Reginald Smith[6] built one-layer the Instant messaging network model. However, in fact most instant messaging tools have groups, such as QQ group, Fetion group and MSN group. Thus, we put forward a two-layer instant messaging network model, as shown in Fig.2.

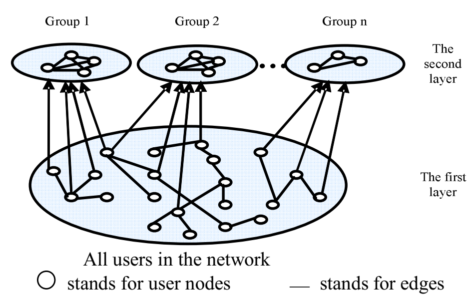

Figure 2 Two-layer model of Instant messaging network

In the two-layer network model, the first layer is used to build the relationship between users and their buddies. All the users are in the network, and the nodes linked by undirected edge can liaise to each other. While the second layer is used to build group network related to users from the first layer, and in which all the user nodes compose a complete graph, that is, all the users can contact each other.

C. Information Propagation Rules

We use the user's status mentioned in literature [8] to build information propagation model in instant messaging network. The network includes N users, and every one may be the one of the three statuses:

Ignorant: They are users having no immune ability to the information propagation. Once they receive the information, they may become spreader, and most users belong to the ignorant in the network.

Spreader: They are the user sending information to the network, and there is small number in the network.

Stifler: They are the users with immune ability to information, and there is small number in the network.

The information propagation rules mentioned in literature [8] consider that there are no stiflers in the network. However, in fact there are some user have immune ability to some information. So, we initialize network with a small part stifler. Secondly, literature [8] ignored the effect of stifler to spreader, but, actually, if users in ones buddy list have immune ability to the information, which means they won't forward information to others, the status of the user may be affected to become stifler. Finally, literature [8] did not consider information propagation in group network. So, we formulate information propagation rules in instant messaging as follows:

Rule1: Suppose nodes i is a spreader. When it meets node j, if j is an ignorant, j becomes a spreader with the probability λ, or becomes a stifler with the probability α; if j is a stifler, the spreader nodes i will become a stifler with the probability α.

Rule2: Suppose nodes i is a spreader, the number of its neighbors is k including w stifler nodes, then the spreader i will become a stifler with the probability $p = w/k$.

Rule3: Suppose nodes i is a spreader in the group network, then nodes i will send information to all user nodes in the group network, and the user nodes will forward information according to Rule1, Rule2, Rule3. Finally, the information propagation goes to be stable, which means no user spread information.

Define $i(t)$, $s(t)$, $r(t)$ are the proportion of ignorant, spreader, stifler in the network respectively, where $i(t) + s(t)+ r(t)=1$ and $i(t) > 0$, $r(t)<1$.

We introduce $Max(S(t))$ and $Final(R(t))$ mentioned in literature[8] to weigh the effect of information propagation, where $Max(S(t))$ stands for the peak value of the amount of spreader in the information propagating, and it reflects the greatest impact caused by information propagation to some extent. $Final(R(t))$ stands for the final value of the amount of stifler while information propagation is over, that is, no spreader is spreading information.

D. *Algorithm Description*

We simulate information propagation in two-layer instant messaging network based on the scale-free network with tunable clustering and the information propagation rules above. The specific algorithm is present as follows.

Step 1 Build the first layer of instant messaging network. The network is consist of N nodes, then, the network is built using scale-free network with tunable algorithm (Section II)

Step 2 Build the second layer of instant messaging network. Initialize group network number and the amount of user nodes in every group network, which are selected randomly from the first layer network, and compose a complete graph.

Step 3 Initialize each node's status. The most are ignorant, the rest are spreader and stifler.

Step 4 Create the integer from 1 to N randomly, the integer stands for the sequence of user nodes to iterate, which make the model are more coincidental to fact.

Step 5 Follow the above information propagation rules and calculate $Max(S(t))$, $Final(R(t))$ and clustering coefficient.

Step 6 Until $S(t)$ reaches to zero, that is no user node spread information.

IV. SIMULATON AND ANALYSIS

We use Matlab7.1 to simulate rumor spreading model (one-layer) and our two-layer information propagation model in this paper respectively, and calculate corresponding clustering coefficient using (2), $Max(S((t))$ and $Final(R(t))$. Then, simulate when all users disconnect the group network. The result is shown from Fig.3 to Fig.10 and Table 1.

A. ***Experiment 1*** *one-layer information propagation model mentioned in literature [8]*

The information propagation rules in literature [8] are as follow. Suppose spread i meets nodes j, if nodes j is ignorant, j will become spreader with the probability λ, and if nodes j is spreader or stifler, nodes j will become stifler with the probability α. It only considered that spreder affects the other nodes.

Firstly, we build the network consist of N user nodes based on scale-free network with tunable clustering, and the probability P_t reaches 0.5, original nodes m_0 reaches 3, edges of new added nodes m reach 3, λ reaches 0.5, α reaches 0.5. Then, follow the above rules to spread information. When network size N reaches 1000, 5000 or 7000, the results of simulation are shown in Fig.3, Fig.4, and Fig.5 respectively.

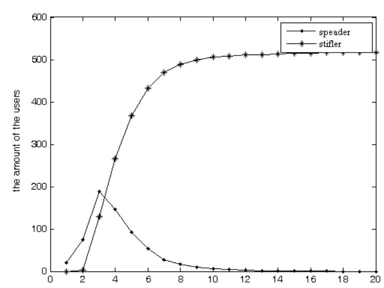

Figure 3 Experiment1 the $S(t)$ and $R(t)$ graph(N=1000)

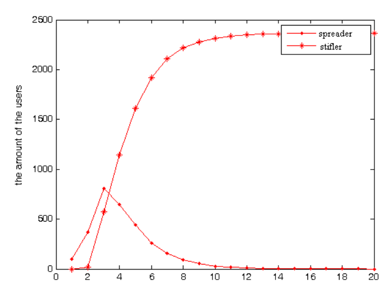

Figure 4 Experiment1 the $S(t)$ and $R(t)$ graph(N=5000)

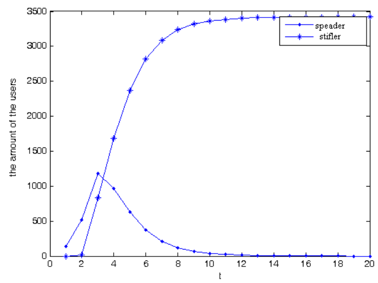

Figure 5 Experiment1 the $S(t)$ and $R(t)$ graph(N=7000)

B. ***Experiment 2*** *two-layer information propagation model in this paper*

(1) When the amount of user nodes in the network N reaches 1000, the number of group network are 10,20,30,40 respectively, and the amount of user nodes in group network is 50, the other parameters are the same as above. The result of simulation is shown in Fig.6.

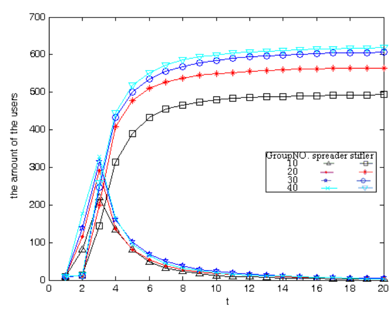

Figure 6 Experiment2 the $S(t)$ and $R(t)$ graph(N=1000)

(2) When the amount of user nodes in the network N reaches 5000, the number of group network are 20,40,60,80 respectively, and the amount of user nodes in group network is 50, the other parameters are the same as above. The result of simulation is shown in Fig.7.

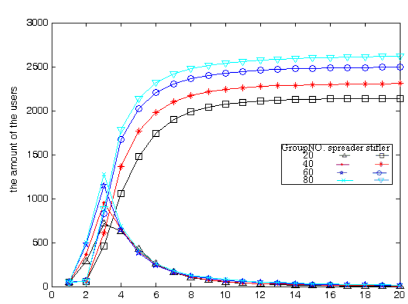

Figure 7 Experiment2 the $S(t)$ and $R(t)$ graph(N=5000)

(3) When the amount of user nodes in the network N reaches 7000, the number of group network are 30,50,70,90 respectively, and the amount of user nodes in group network is 50, the other parameters are the same as above. The result of simulation is shown in Fig.8.

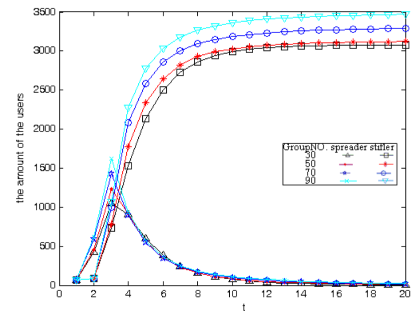

Figure 8 Experiment2 the $S(t)$ and $R(t)$ graph(N=7000)

Fig.3 to Fig.8 shows that whatever N reaches1000, 5000, or 7000, all spreaders will disappear, and this is corresponding with fact. As we can see from Table 1, in the case of having group network, the more the group number is, the larger the peak value of spreaders and the final value of stifler are. The clustering coefficient of the network with group network is larger than that of the network with no group network. Sometimes, the clustering coefficient of the network does not increase as the amount of group network increase. That is because the user may be in many group networks at the same time. Generally speaking, the clustering coefficient of the network is larger as the amount of group network increases.

TABLE 1 THE DATA OF TWO SIMULATIONS

N		**Group NO.**	***C***	***Max (S(t))***	***Final (R(t))***
1000	Experiment1	0	0.2775	189	516
	Experiment2	10	0.2953	220	494
		20	0.3032	292	565
		30	0.3015	315	606
		40	0.2888	329	619
5000	Experiment1	0	0.2754	807	2366
	Experiment2	20	0.2823	716	2140
		40	0.2859	954	2307
		60	0.2909	1147	2495
		80	0.2872	1277	2619
7000	Experiment1	0	0.2730	1183	3419
	Experiment2	30	0.2798	1046	3080
		50	0.2822	1236	3122
		70	0.2824	1423	3219
		80	0.2838	1626	3473

C. ***Experiment 3*** *break the link between user and group network in two-layer information propagation model*

In order to test the effect of group network to propagate information, now we make the information propagation simulation with group network. Moreover, we make the simulation with all the users in the network disconnect with group network(s).

(1) When the network size N reaches 1000, the number of group network reaches 80, and the other parameters are the same as above. The result of $S(t)$ and $R(t)$ is shown in Fig.9.

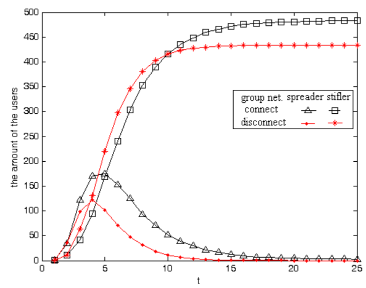

Figure 9 Experiment 3 $S(t)$ and $R(t)$ graph with connecting and disconnecting group network (N=1000)

(2) When the network size N reaches 5000, the number of group network reaches 100, and the other parameters are the same as above. The result of $S(t)$ and $R(t)$ is shown in Fig.10.

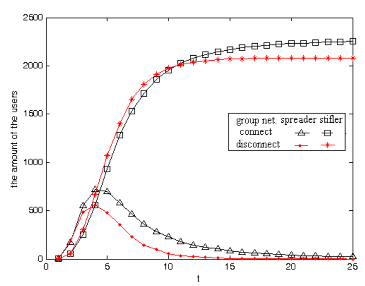

Figure 10 Experiment 3 $S(t)$ and $R(t)$ graph with connecting and disconnecting group network (N=5000)

From Fig.9 to Fig.10, as the time goes on and the spreader and the stifler effect each other, when users connect with group network(s), the peak value of the amount of the spreader ($Max(S(t))$) and the final value of the amount of the stifler ($Final(R(t))$) are larger than those when all the users disconnect with group networks.

From Experiment 1 to Experiment 3, we find that due to the existence of the group network, the users contact closer, the information propagate more efficiently, for $Max(S(t))$ and $Final(R(t)$ are larger.

V. CONCLUSIONS

In this paper, we have proposed a new two-layer model of Instant Messaging Network by use of scale-free networks with tunable clustering and improve on the information propagation mentioned in literature [8]. By simulation and experiment, we have observed that the diversification of spreader and stifler. Finally all spreaders disappear. And there are only stifler and ignorant exit, which is corresponding with the fact. The results of simulation show that the group network plays a vital role of information propagation. Suppose we would like to effect information propagation efficiently, we might break some links between users and the group network(s) to stop the information propagation in some group network.

In this paper we only take account into the clustering coefficient to affect the information propagation in instant messaging network. There are other factors required us to further study, such as the sensibility of the user in instant messaging network, the delay of information to propagate and so on. Moreover, we only make statistic to the peak value of spreader and the final value of stifler. In the future, we hope to find a better measure of the scope of information propagation.

ACKNOWLEDGMENTS

We would like to thank the many team members including schoolmates and teachers who help this research be possible. Then, this work is supported by the follow grants from National Basic Research Program of China (973Program) (No.2008CB317111), National Natural Science Foundation of China (No.60873079), Program for New Century Excellent Talents in University of China and Natural Science Foundation of Chongqing of China (No.2008BB2241).

REFERENCES

[1] D J Watts, S H Strogatz, "Collective dynamics of 'small-world' network", Nature, London, Vol. 393, pp. 440-442, June 1998.

[2] A L Barabasi, R Albert, "Emergence of scaling in random networks", Science, New York, Vol. 286, pp. 509-511, October 1999.

[3] A L Barabasi, R Albert, H Jeong, "Mean-field theory for scale-free random networks", Physic A., Amsterdam, Vol. 272, pp.173-187, 1999.

[4] Liu Changyu, Hu Xiaofeng,Si Guangya,Luo Pi, "Public Opinion Propagation Model Based on Small World Networks". Journal of System Simulation, China, Vol. 18, pp. 3608-3610, 2006.

[5] Yamir Moreno, Maziar Nekovee, Amalio F. Pachecol, "Dynamics of rumor spreading in complex networks", Phys.Rev.E, New York, Vol. 69, pp.066130-066136,2004.

[6] Reginald Smith, "Instant Messaging as a scale-free network", eprint arXiv: cond-mat/0206378 June 2005, http://arxiv.org/abs/cond-mat/0206378 [Accessed: June 20, 2005].

[7] P Holme, B J Kim, "Growing scale-free networks with tunable clustering", Phys.Rev.E, Melville, Vol. 65, pp. 026107-026109, 2002.

[8] Pan Zaofeng, Wang Xiaofan, Li Xiang, "Simulation Investigation on Rumor Spreading on Scale-free Network with Tunable Clustering", Journal of System Simulation, China, Vol. 18, pp. 2346-2348, 2006.