

Research Article

Research on Effectiveness Modeling of the Online Chat Group

Hua-Fei Zhang,¹ Li-Gang Dong,¹ Jia-wei Sun,² and Ying Li¹

¹ School of Information & Electronic Engineering, Zhejiang Gongshang University, Hangzhou, 310018, China

² Information Management Center, Training Department, Information Engineering University, Zhengzhou, 450002, China

Correspondence should be addressed to Li-Gang Dong; donglg@mail.zjgsu.edu.cn

Received 25 May 2013; Revised 9 September 2013; Accepted 10 September 2013

Academic Editor: Panos Liatsis

Copyright © 2013 Hua-Fei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The online chat group is a small-scale multiuser social networking platform, in which users participate in the discussions and send and receive information. Online chat group service providers are concerned about the number of active members because more active members means more advertising revenues. For the group owners and members, efficiency of information acquisition is the concern. So it is of great value to model these two indicators' impacting factors. This paper deduces the mathematical models of the number of active members and efficiency of information acquisition and then conducts numerical experiment. The experimental results provide evidences about how to improve the number of active members and efficiency of information acquisition.

1. Introduction

With the development of online social networks, a kind of small-scale platform, the online chat group, appears. Online chat group is an extension of the social group on online social networks. Therefore, they have similarities. For example, scale of online chat group accords with “Dunbar’s Number” [1]; it gathers people by region, interest, or task, and there is a certain degree of social interaction or social relation among group members.

A central theme of researches on group is group effectiveness, which has been studied in the literatures. For instance, Reagans and Zuckerman’s research [2] claimed that high network density improves group effectiveness by enhancing coordination and trust among their members; Oh et al. [3] analyzed such influencing factors as group closure, bridging relationships, and current and past relationships on the social capital. They argued that greater group social capital leads to greater group effectiveness. As to the realistic interaction group, similar to online chat group, Dong et al. [4] discovered that higher turn-taking rate and backchannel rate increase the group effectiveness while higher turn competition rate and uneven turn transferring decrease the group effectiveness. Obviously, the yardsticks are as varied as the type of groups.

And distinctions exist between online chat groups and other types of groups. In contrast to groups in the physical offline world, online chat groups are organized in the virtual world, other than the form of sitting face-to-face. And some advantages are listed as follows [5]. (1) Members with various backgrounds can communicate without the geographical restrictions. (2) It is easy to communicate for information. (3) It eases the process of archiving and retrieving historical chats. (4) Chats are in public form and everybody gets the equality of speaking. Moreover, online chat group service providers are concerned about the number of active members because more active members means more advertising revenues. For the group owners and members, the information acquisition is the concern, whose importance for members of online groups has been underlined in literatures [5–7] as well. Finally, the online chat groups we are studying are the synchronous ones, while most related work about online groups, like Usenet newsgroups [8], Yahoo groups [9], Google groups [10], and so forth, referred to the asynchronous manner.

From the perspective of means, most studies of online group mainly focus on information mining, including group detection and topic detection. For group detection, Mutton [11], utilizing the timing characteristics of the chat data, proposed three heuristic rules to tap the communication

relationship between users; Acar et al. [12] found smaller discussion groups from the chat room by tensor (user \times keyword \times time) analysis techniques, and in a recent study, an approach called multilayered edge clustering coefficient to detect groups from multilayered social network is developed [13]. As for topic detection, most technologies are text-based analysis such as Butterfly system [14] and ChatTrack system [15]. In fact, question of topic detection in group is nearly the same as that in other text-based environment. So the means in [16, 17] are suitable for online chat group as well.

In general, these studies employ methods based on user-generated data combined with data mining techniques, to develop some algorithms or systems. They are the basis of application. However, they do not directly reveal any related laws or characteristics of the effectiveness of the online chat group.

So this paper studies the effectiveness of online chat groups, and it proceeds as follows. Firstly, we give a description of the online chat group and make some assumptions. Secondly, we model group members' information acquisition efficiency and the number of active members. At last, we conduct the numerical experiment to reveal the effect of the related parameters on the online chat group.

2. Basic Concepts, Assumptions, and Symbol Definitions

Because users join the newly created group with a purpose of participating in the group chat, all the group members are assumed to be active at the beginning. Each member has a certain number of topics of interest. As time goes, some members may be annoyed with the continuous useless topics and thus shield (i.e., ignore) messages from the group and become inactive. However, inactive members may convert from inactive state to active state after a length of time. Thus, every member is in either inactive or active state.

Before deducing the mathematical models, we give the following assumptions.

Assumption 1. The total number of online members (including active members and inactive members) in a group keeps constant for the duration when we analyze the group. In other words, neither new members join the group nor old members drop out.

Assumption 2. At any moment, the group has only one topic being discussed. That is, one topic must be stopped before another begins.

Assumption 3. Each topic appears with the same probability.

Assumption 4. Either active or inactive state duration of each member per cycle, experiencing two complete states as a cycle, is subjected to the exponential distribution.

Definitions of symbols used in the model are shown in the Tables 1 and 2.

TABLE 1: Symbols of independent variables.

Symbol	Definition
N	Number of group members
m_i	Number of topics that the i th member is interested in
r_i	Number of uninteresting topics that the i th member can tolerate
$1/\beta$	Average duration of each group member's inactive state per cycle
$1/\delta$	Average duration of each topic
w	Average number of repeated interesting topics per member

TABLE 2: Symbols of dependent variables.

Symbol	Definition
N_a	Number of active members in a steady group
m	Average number of interesting topics per member
M	Number of topics in a group
$1/\alpha$	Average duration of each group member's active state per cycle
r	Average number of uninteresting topics that each member can tolerate
P	Appearance probability of interesting topics
s	Average number of interesting topics acquired within unit time by a member per cycle
s_i	Average number of interesting topics acquired by a member per cycle
s_u	Average number of uninteresting topics acquired by a member per cycle
μ	The ratio of interesting topics acquired by a member

3. Model of Information Acquisition Efficiency

Number of topics that the i th member is interested in is m_i ; so, average number of interesting topics per member is

$$m = \frac{\sum_{i=1}^N m_i}{N}. \quad (1)$$

Between any two members, they may have a common part of topics which they are interested in. For example, member A is interested in football and basketball, while member B is interested in table tennis and basketball. Thus, the repeated interesting topic to them is basketball.

Topics in a group are the union set of topics of interest to every member. In our model, the average number of repeated topics of interest to each member is expressed as w . N members group has $(N-1)w$ repeated topics. So, excluding repeated interesting topics of each member, total number of topics in a group is

$$M = \begin{cases} Nm - (N-1)w, & N \geq 1 \\ 0, & N = 0. \end{cases} \quad (2)$$

According to Assumption 3, the appearance probability of each topic is $1/M$. So, to a member, the ratio of interesting topics among all topics is equivalent to the appearance probability of interesting topics. Consider

$$P = \frac{m}{M} = \frac{m}{Nm - (N-1)w}. \quad (3)$$

Average duration of each topic obeys exponential distribution. Assuming its parameter is δ , average duration of each topic is $1/\delta$.

Similar to the average number of interesting topics per member, average number of uninteresting topics that each member can tolerate is

$$r = \frac{\sum_{i=1}^N r_i}{N}. \quad (4)$$

It is easy to derive the number of topics experienced by a member throughout one cycle, during which there are m_x topics. During the cycle, any topic can be repeatedly discussed. The cycle will not be finished unless the number of uninteresting topics is more than r_i . Therefore, m_x can be positive infinity. In detail, the range of m_x is $[r, +\infty)$. The probability distributions of the number of topics experienced by a member throughout one cycle are shown as follows:

$$\begin{aligned} P_r &= P(m_x = r) = C_{r-1}^0 (1-P)^r, \\ P_{r+1} &= P(m_x = r+1) = C_r^1 P(1-P)^r, \\ P_{r+2} &= P(m_x = r+2) = C_{r+1}^2 P^2(1-P)^r, \\ &\vdots \\ P_k &= P(m_x = k) = C_{k-1}^{k-r} P^{k-r} (1-P)^r \\ &\vdots \end{aligned} \quad (5)$$

As shown above, average duration of each topic is $1/\delta$. If a member experiences i topics during one active state, the duration of the active state is $i \cdot 1/\delta$, and the number of interesting topics experienced is $i - r$. Therefore, average number of interesting topics acquired within unit time by a member per cycle is

$$s = \sum_{i=r}^{+\infty} P_i \frac{i-r}{1/\delta}. \quad (6)$$

Then average number of interesting topics acquired by a member per cycle and uninteresting topics acquired by a member per cycle are

$$\begin{aligned} s_i &= \sum_{i=r}^{+\infty} P_i (i-r), \\ s_u &= r. \end{aligned} \quad (7)$$

Finally, information acquisition efficiency can be defined as the ratio of interesting topics among topics acquired by

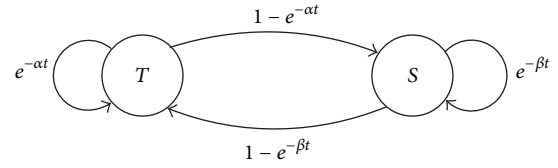


FIGURE 1: State transition diagram of each member.

a member, that is, average number of interesting topics acquired by a member per cycle / (average number of interesting topics acquired by a member per cycle + average number of uninteresting topics acquired by a member per cycle); namely,

$$\mu = \frac{s_i}{s_i + s_u}. \quad (8)$$

4. Model of Number of Active Members

In our model, average duration of each member's active state per cycle is

$$\frac{1}{\alpha} = \sum_{i=r}^{+\infty} P_i \cdot i \frac{1}{\delta}. \quad (9)$$

According to Assumption 4, either active or inactive state duration of each member per cycle follows exponential distribution, parameters of which are α and β , respectively. Then the active state probability of each member is $e^{-\alpha t}$, and inactive state probability is $e^{-\beta t}$. Let active state be represented by T , and let inactive state be represented by S . Its state changing process can be described by the Ehrenfest model [18], a birth-death process to explain the second law of thermodynamics. The model considers N particles in two containers. Particles independently change container at a rate λ . We consider particles to be the members of a group, and two containers represent two states. But in this case, birth rate and death rate are not necessarily equal. Thus, state transition diagram of each member can be shown as Figure 1.

And stationary distribution of active member number follows binomial distribution, $B(N, \beta/(\alpha + \beta))$ [19].

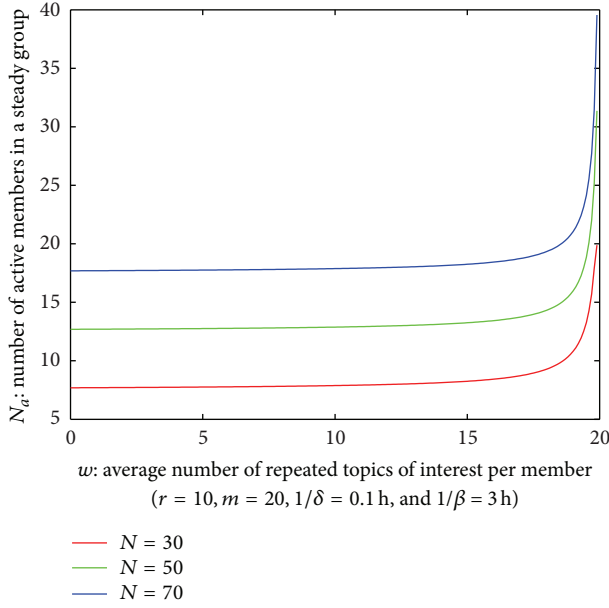
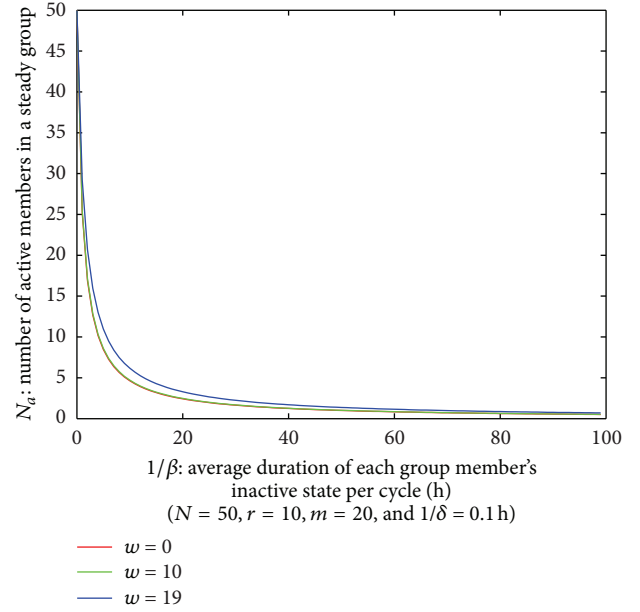
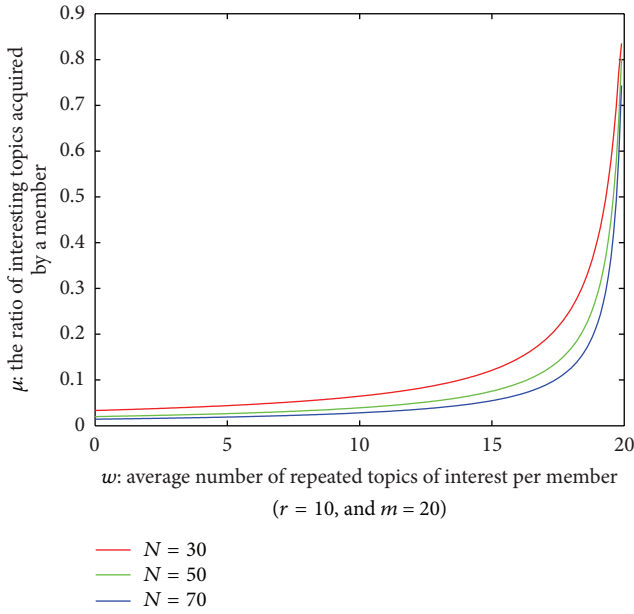
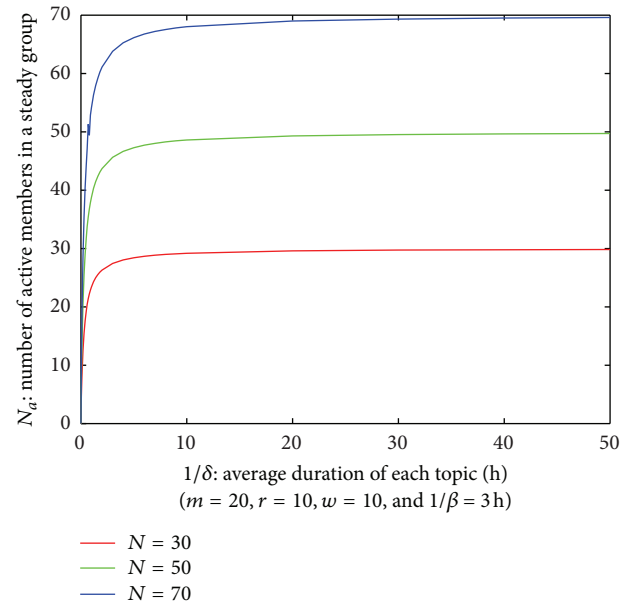
Therefore, when the group is steady, the expected number of active members is

$$N_a = \frac{N\beta}{\alpha + \beta}. \quad (10)$$

5. Numerical Experiments and Results Analysis

We carried out comprehensive numerical experiments to disclose the effect on the number of active members and efficiency of information acquisition.

- (a) Let us set $1/\delta = 0.1$ hr, $1/\beta = 3$ hr, $r = 10$, and $m = 20$. Figures 2 and 3 show the relationships of N_a versus w and μ versus w , respectively.

FIGURE 2: N_a versus w .FIGURE 4: N_a versus $1/\beta$.FIGURE 3: μ versus w .FIGURE 5: N_a versus $1/\delta$.

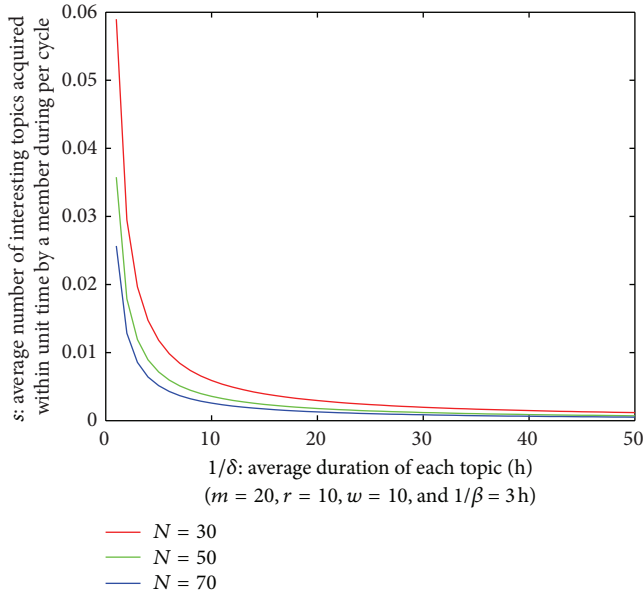
When the average number of repeated interesting topics per member is small, both the number of active members in a steady group and the ratio of interesting topics acquired by a member are less affected. When the average number of repeated interesting topics per member is close to the average number of interesting topics per member, it greatly affects the number of active members of a group and the ratio of interesting topics acquired by a member.

- (b) Let us set $N = 50, r = 10, m = 20$, and $1/\delta = 0.1$ hr. Figure 4 shows the relationship of N_a versus $1/\beta$.

With the increase of the average duration of each group member's inactive state per cycle, the number of active members in a group becomes less (as shown in Figure 4).

The ratio of interesting topics acquired by a member is unrelated to the average duration of each group member's inactive state per cycle.

- (c) Let us set $m = 20, r = 10, w = 10$, and $1/\beta = 3$ hr. Figures 5 and 6 show the relationships of N_a versus $1/\delta$ and s versus $1/\delta$, respectively.

FIGURE 6: s versus $1/\delta$.

Longer average duration of each topic means less topics that the group members are not interested in. Then there will be more group members staying in active state. If the average duration of each topic was long enough, all the group members are active (as shown in Figure 5).

The ratio of interesting topics acquired by a member is unrelated to the average duration of each topic.

Longer average duration of each topic means less interesting topics acquired within unit time by a member per cycle (as shown in Figure 6).

- (d) Let us set $1/\beta = 3$ hr, $m = 20$, $1/\delta = 0.1$ hr, and $w = 10$. Figures 7 and 8 show the relationships of N_a versus r and s versus r , respectively.

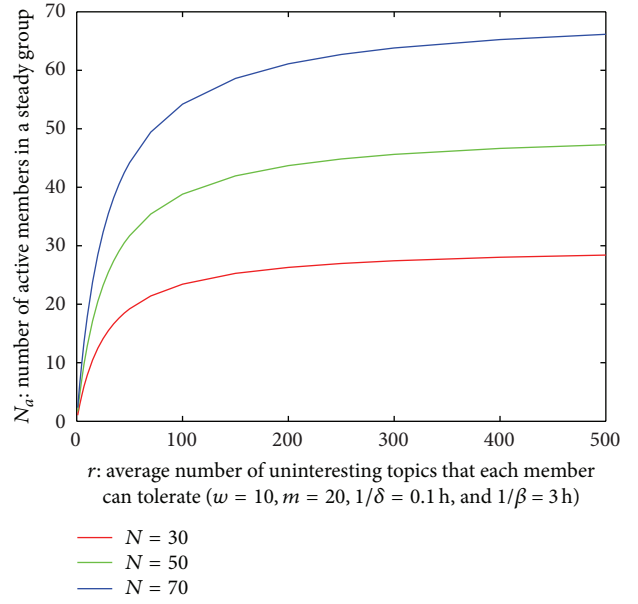
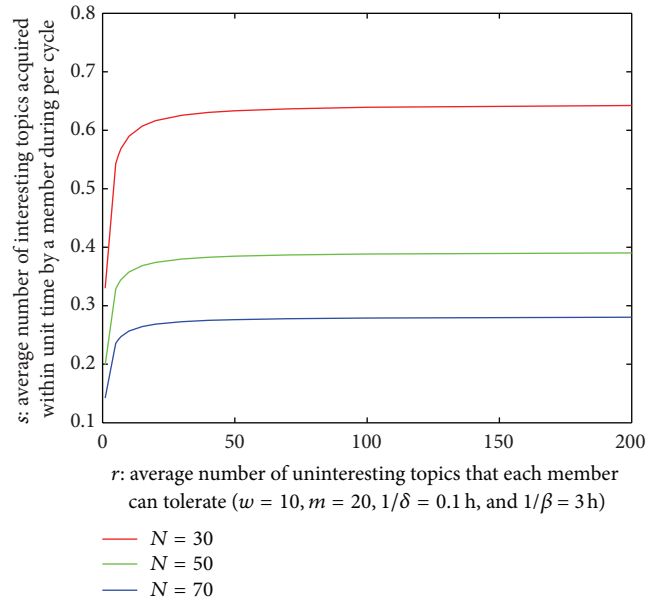
More uninteresting topics that each member can tolerate means longer inactive time of the group members. Then there will be more group members staying in the active state. If the average number of uninteresting topics that each member can tolerate was big enough, all the group members are active (as shown in Figure 7).

The ratio of interesting topics acquired by a member is unrelated to the average number of uninteresting topics that each member can tolerate.

The variation of the average number of interesting topics acquired within unit time by a member per cycle has the similar trend with the number of active members in a group (as shown in Figure 8).

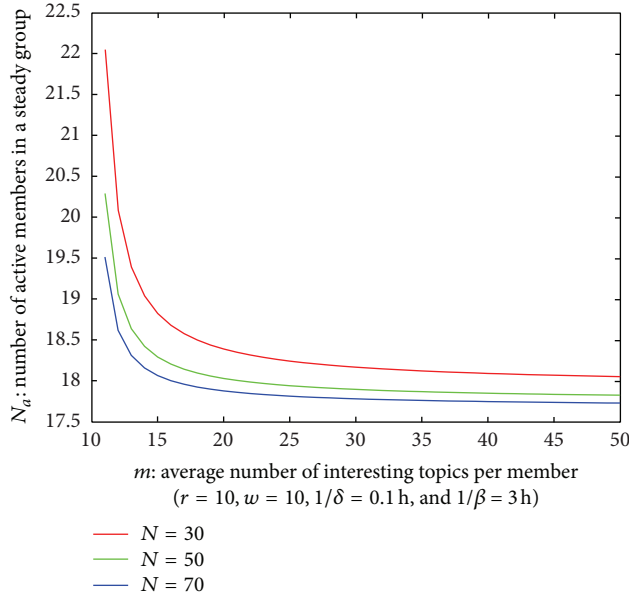
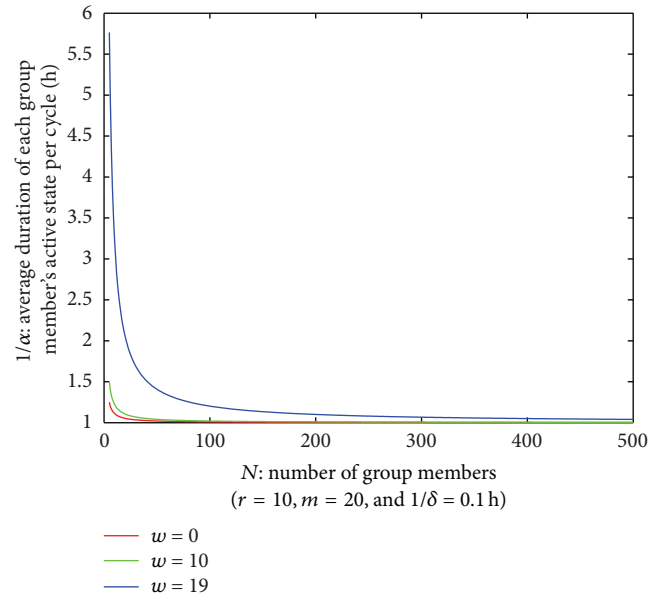
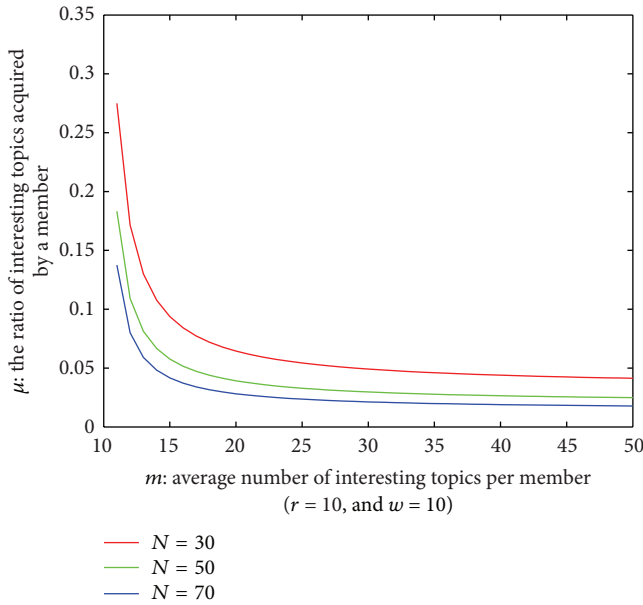
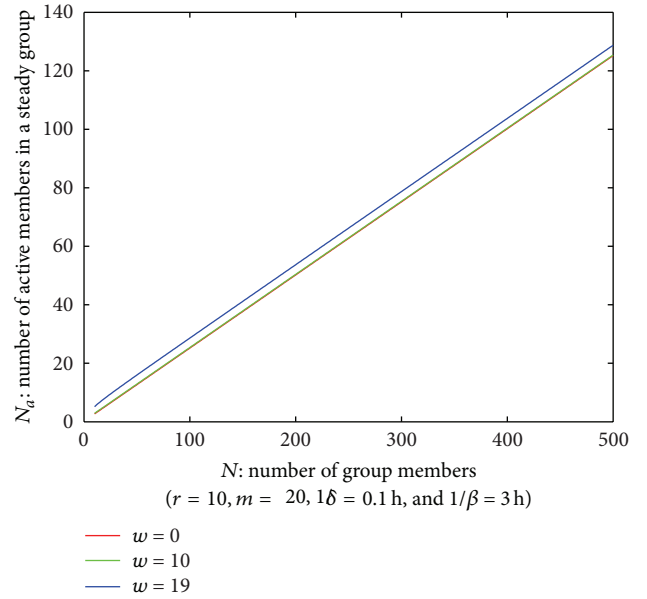
- (e) Let us set $1/\beta = 3$ hr, $w = 10$, $1/\delta = 0.1$ hr, and $r = 10$. Figures 9 and 10 show the relationships of N_a versus m and μ versus m , respectively.

More interesting topics per member (while the number of the repeated topics is kept the same) means more diversity of interesting topics among members. Then the number of uninteresting topics will increase. It will cause that fewer

FIGURE 7: N_a versus r .FIGURE 8: s versus r .

group members remain in the active state. The ratio of interesting topics acquired by a member will also decrease. When the average number of interesting topics per member is large enough, the number of active members in a group and the ratio of interesting topics acquired by a member will be less affected.

The variation of the average number of interesting topics acquired within unit time by a member per cycle is the same with the ratio of interesting topics acquired by a member.

FIGURE 9: N_a versus m .FIGURE 11: $1/\alpha$ versus N .FIGURE 10: μ versus m .FIGURE 12: N_a versus N .

(f) Let's set $1/\beta = 3$, $m = 20$, $1/\delta = 0.1$, $r = 10$. Figures 11, 12 and 13 show the relationships of $1/\alpha$ versus N , N_a versus N and μ versus N , respectively.

More group members means more active members in a group. More group members will give more topics and increase the ratio of uninteresting topics among all topics. That will make the average duration of each group member's active state shorter (as shown in Figure 11).

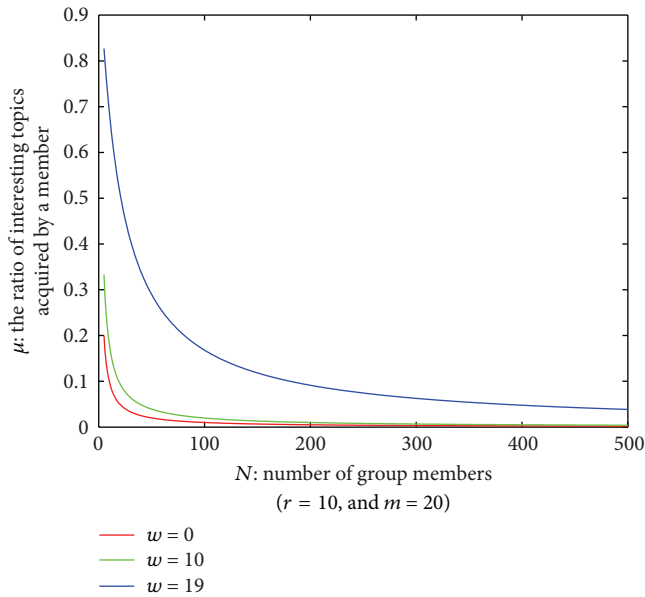
Figure 12 shows that the number of active members of a group and the number of group members have incremental linear relationship.

To each group member, the number of his interesting topics being discussed is reduced. So the ratio of interesting topics acquired by a member is smaller (as shown in Figure 13).

The variation of the average number of interesting topics acquired within unit time by a member per cycle is the same with the ratio of interesting topics acquired by a member.

6. Conclusions

The main contribution of this paper is modeling of the number of active group members and efficiency of information

FIGURE 13: μ versus N .

acquisition. According to the models and their numerical experiments results, we have the following conclusions. (1) In the case that other conditions remain unchanged, more group members means more active members, but there will be a sharp decline in duration of each group member's active state and members' efficiency of information acquisition. Effectiveness of the group deteriorates instead, so it is not so good that too many members stay in a group. (2) For group service providers, to shorten the inactive duration can significantly improve the effectiveness of the group. Therefore, we recommend that the group service providers should routinely publish attractive contents to attract inactive members to return to the active state. (3) In order to ensure that the group has a higher activity, the group administrators should absorb members with more common topic. (4) In order to get more information, group members should improve their tolerance. As a result, group will be more active. (5) Too many topics of interest will lead the proposed topics to be more dispersed so that some group members lose interest in the group. Therefore, we recommend that the interest scope of a group should be set narrow.

This study may provide a reference value for the development of the group's management and related studies. Besides, this study is only for the homogenous case. Namely, we assume that each group member has the same parameter or average value. Heterogeneous case is our future study.

Acknowledgments

This work was supported by a grant from the National Natural Science Foundation of China (61170215 and 61102074).

References

- [1] B. Gonçalves, N. Perra, and A. Vespignani, "Modeling users' activity on Twitter networks: validation of Dunbar's number," <http://arxiv.org/pdf/1105.5170v2.pdf>.
- [2] R. Reagans and E. W. Zuckerman, "Networks, diversity, and productivity: the social capital of corporate R&D teams," *Organization Science*, vol. 12, no. 4, pp. 502–517, 2001.
- [3] H. Oh, G. Labianca, and M.-H. Chung, "A multilevel model of group social capital," *Academy of Management Review*, vol. 31, no. 3, pp. 569–582, 2006.
- [4] W. Dong, B. Lepri, T. Kim, F. Pianesi, and A. P. Pentland, "Modeling conversational dynamics and performance in a social dilemma task," in *Proceedings of the 5th International Symposium on Communications, Control and Signal Processing*, pp. 1–4, Rome, Italy, 2012.
- [5] C. Budak and R. Agrawal, "On participation in group chats on twitter," in *Proceedings of the International World Wide Web Conference*, Rio, Brazil, 2013.
- [6] P. J. Bateman, P. H. Gray, and B. S. Butler, "Community commitment: how affect, obligation, and necessity drive online behaviors," in *Proceedings of the 6th International Conference on Intelligent Computing and Intelligent Systems*, p. 63, Madison, Wis, USA, 2006.
- [7] P. B. Brandtzæg and J. Heim, "User loyalty and online communities: why members of online communities are not faithful," in *Proceedings of the 2nd International Conference on Intelligent Technologies for Interactive Entertainment*, Cancun, Mexico, 2008.
- [8] J. Arguello, B. Butler, E. Joyce, R. Kraut, K. S. Ling, and X. Wang, "Talk to me: foundations for successful individual-group interactions in online communities," in *Proceedings of the Conference on Computer-Human Interaction*, pp. 959–968, Montreal, Canada, April 2006.
- [9] L. Backstrom, R. Kumar, C. Marlow, J. Novak, and A. Tomkins, "Preferential behavior in online groups," in *Proceedings of the International Conference on Web Search and Data Mining (WSDM '08)*, pp. 117–128, Stanford, Calif, USA, February 2008.
- [10] D. A. Huffaker, C. Teng, P. S. Matthew, L. Gong, and L. A. Adamic, "The impact of group attributes on communication activity and shared language in online communities," *First Monday*, vol. 16, no. 4, 2011.
- [11] P. Mutton, "Inferring and visualizing social networks on internet relay chat," in *Proceedings of the 8th International Conference on Information Visualisation (IV '04)*, pp. 35–43, London, UK, July 2004.
- [12] E. Acar, S. A. C. Amtepe, M. S. Krishnamoorthy, and B. Yener, "Modeling and multiway analysis of chatroom tensors," in *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, pp. 256–268, Atlanta, Ga, USA, 2005.
- [13] P. Bródka, T. Filipowski, and P. Kazienko, "An introduction to community detection in multi-layered social network," *Information Systems, E-Learning, and Knowledge Management Research*, vol. 278, pp. 185–190, 2013.
- [14] N. W. van Dyke, H. Lieberman, and P. Maes, "Butterfly: a conversation-finding agent for Internet relay chat," in *Proceedings of the 4th International Conference on Intelligent User Interfaces (UIST '99)*, pp. 39–41, Los Angeles, Calif, USA, November 1998.
- [15] J. Bengel, S. Gauch, E. Mittur, and R. Vijayaraghavan, "Chat-Track: chat room topic detection using text classification," in

Proceedings of the 2nd Symposium on Intelligence and Security Informatics, pp. 266–277, Tucson, Ariz, USA, 2004.

- [16] X. Guo, Y. Xiang, Q. Chen, Z. Huang, and Y. Hao, “LDA-based online topic detection using tensor factorization,” *Journal of Information Science*, vol. 39, no. 2, pp. 1–11, 2013.
- [17] S. Liu and H. Chang, “A topic detection and tracking system with TF-Density,” *Recent Progress in Data Engineering and Internet Technology*, pp. 115–120, 2013.
- [18] S. Karlin and J. McGregor, “Ehrenfest urn models,” *Journal of Applied Probability*, vol. 2, no. 2, pp. 352–376, 1965.
- [19] O. Krat and M. Schaefer, “Mean passage times for tridiagonal transition matrices and a two-parameter Ehrenfest urn model,” *Journal of Applied Probability*, vol. 30, no. 4, pp. 964–970, 1993.

