Evaluation of Characteristics of a Community for Recommending Dissimilar Communities in SNS

SUSUMU TAKEUCHI, MASANORI AKIYOSHI, NORIHISA KOMODA
Osaka University
Graduate School of Information Science and Technology
Yamadaoka 2-1, Suita, Osaka 565-0871
JAPAN

Abstract: Currently, SNSes (Social Networking Services) are widely available on the Internet. In a SNS, users can communicate with the users who have a similar interest in a community. To activate communications drastically in SNS, users are encouraged to join other dissimilar communities. In this paper, utilizing activities of communities and relationships with communities is proposed to realize a dissimilar community recommendation. However, such communities tend to be useless for the users, so that investigating characteristics of communities that are selected by the methods is necessary to recommend. By utilizing real data in the largest SNS in Japan, the correlations between the users' subjective estimations and the characteristics of communities are evaluated. As a result, it is clarified that effective recommendation of dissimilar communities will be possible by integrating activities of communities and relationships with communities.

Key–Words: SNS, Social network, Activation, Community, Recommendation

1 Introduction

Recently, SNSes (Social Networking Services) such as Orkut [1] and LinkedIn [2] are widely available on the Internet. In those services, a user registers well-known people as 'friends' in the services, and communicates with other users, which are not limited to the friends. The user may communicate with unknown users as well as traditional communication systems such as anonymous BBS, but the user has *indirect connections* with the other unknown users on the social network so that communication in SNSes is considered more reliable than the traditional systems. Therefore, SNSes that utilize social network are expected to be a useful inter-personal communication support system in daily life [3].

Ordinary SNSes provide several communication tools which encourage the users to communicate, e.g., messaging, chat, blogging, communities, etc. For activating communications in a SNS, this paper focuses on communities, i.e., forums or discussion groups, because other tools such as messaging, chat, and blogging mainly encourages the users to communicate with known people. Communities are provided for users to communicate with other users about specific topics such as hobbies, jobs, fashion, and politics by using BBS. However, since a community is a closed space, information exchanged in the community does not propagate easily to others. Thus, once activities in communities weaken, communications of the entire service decrease so that maintaining the service will become difficult.

To solve this issue, recommending communities to users is considered necessary. The users who are recommended communities may alter their behaviors, so communications in SNS may be varied and activated. Although recommending similar communities cannot encourage the users fundamental change of their behaviors or interests, continuous recommendation of *dissimilar communities* that are not similar to the existing interests of the users is considered to be required for activating communications. However, because those communities do not relate to the users' interests, most of recommended communities are needless for the users. Thus, dissimilar communities must be selected and recommended by an appropriate method.

Therefore, the methodology for recommending dissimilar communities, which utilizes activities of communities and social relationships with the members in a community, is proposed in this paper. Since the correlation between the characteristics of a community that can be obtained by the proposed methods and users' subjective estimations is not clear, tendencies of the characteristics should be investigated. Consequently, by performing an experiment with the real users and communities in a SNS, this paper unveils

what kind of characteristics of a community should be considered to recommend appropriate dissimilar communities.

The rest of the paper is organized as follows. In Section 2, the outline of SNSes and the necessity of recommending communities are described. Section 3 presents a methodology to recommend dissimilar communities, and Section 4 describes the details of the experiment and evaluations. Finally, Section 5 concludes the paper.

2 Activating Communications in SNS

2.1 Overview of a SNS

Most SNSes provide a collection of communication tools which encourage the users to communicate, e.g., messaging, chat, blogging, communities, etc. SNSes are different from existing communication support systems such as BBS in *non-anonymity*, so communications in SNSes are believed to be more reliable even if a user communicates with unknown users.

In ordinary SNSes, this non-anonymity is guaranteed by forming a social network. A general social network is constructed by social relationships in the real world, but the users in SNSes sometimes list unknown users so that social networks in SNSes cannot be equivalent or sub graphs of the social network in the real world. This means that a user does not always know well about the user's *buddies* in the SNS compared to the friends in the real world.

By utilizing non-anonymity of the services, reliable communities are provided to communicate about specific topics, e.g., hobbies, jobs, fashion, and politics. In the community, a user can discuss the topic or obtain information related to the topic in which the user is interested. Thus, users in SNSes can communicate not only with known familiar users, but with unknown users who have same interest at the same service.

2.2 Issues of Activating Communications in Existing SNSes

SNSes provide varieties of communication support methods for enabling the users to communicate for various purposes. However, current services cannot provide any means to activate communications drastically.

First, email, messaging, and chat are provided to communicate with limited users, particularly friends. When a user wants to communicate with other users, the user should find them by other methods. Thus, these methods can activate between fixed users, but cannot activate communications among the users fundamentally. Secondly, blogging is provided to publish users' comments. When a user searches a specific topic or follows links of blogs, the user can refer new information so that communications in SNSes are possible to be varied. However, such actions of the users are necessary to activate communications on the blogs so that activating halfhearted users is difficult. Lastly, communities are provided to communicate about specific topics with several kinds of the users. Thus, the members of a community can discuss with many users and can get new information, which relates to the users' interests. If all of the communities in SNS is thriving and all of the users joins some of them, communities are possible to encourage the users to communicate actively. However, activity of a community is different according to the nature of the community and is varied with time. In addition, not all the users join thriving communities or alter participated communities for maintaining their activities.

Therefore, only providing messaging, chat, blogging, or communities are not enough to activate communications in SNS, especially for not active users.

2.3 Recommending Communities

To activate communications in a SNS, recommending communities to the users is considered one of effective methods. The users who refer recommended communities may join some of them, so that recommending communities encourages the users to change their habits or behaviors fundamentally.

The methods of recommending communities can be considered to be categorized into the following three means.

2.3.1 Recommending Popular Communities

Recommending popular communities, i.e., thriving or huge communities, to users is effective especially for halfhearted users, because the users are expected to get interested in such communities where many users communicate or many messages are exchanged. Thus, popular communities will help to give new perspectives to the users.

However, users do not always get interested in popular communities. On the contrary, they sometimes get interested in *minor* communities, which do not have many users and exchanges of information. Therefore, recommending only popular communities is not enough to activate communications fundamentally in a SNS.

2.3.2 Recommending Similar Communities

Communities which are similar to the communities a user has already joined can be candidates for a part of the communities in which the user may get interested. Spertus et al. [4] propose multiple methods to measure similarity of communities in Orkut. Those methods will enable the users to find similar communities and to join them.

However, since such similar communities are similar to the existing habits or behaviors of the users, similar communities can affect only a part of their perspectives. Thus, fundamental change of them will take long time so that activating communications drastically is difficult.

2.3.3 Recommending Dissimilar Communities

Recommending dissimilar communities to users, i.e., recommending communities that are not similar to the communities the users have already joined, is considered effective to affect the users' habits or behaviors fundamentally. The users who are recommended dissimilar communities will refer and join some of them. Thus, dissimilar communities can be considered to give new perspectives to the users, and to help to activate communications drastically.

However, since these communities do not relate to the users' interests, most of the recommended communities are useless for the users. Therefore, determining what kind of dissimilar communities should be recommended is difficult.

Note that almost of the popular communities a user does not join are dissimilar communities for the user. Hence, "dissimilar communities" include "popular communities" in this paper.

3 Methodology for Recommending Dissimilar Communities

3.1 Overview

As described in Section 2.3.3, recommending dissimilar communities, which include popular communities explained in Section 2.3.1, is effective to activate communications fundamentally in a SNS. However, those communities tend to be needless for the users, so recommending the communities that are dissimilar to the users' interests but may fascinate the users is necessary.

Therefore, focusing on the activities of communities and relationships with the members of communities is proposed to utilize for determining appropriate dissimilar communities in this research. By integrating both of them, various kinds of appropriate dissimilar communities will be recommended to the users. Figure 1 illustrates the overview of the proposed methodology for recommending dissimilar communities. The two methods are explained in details in the following sections.

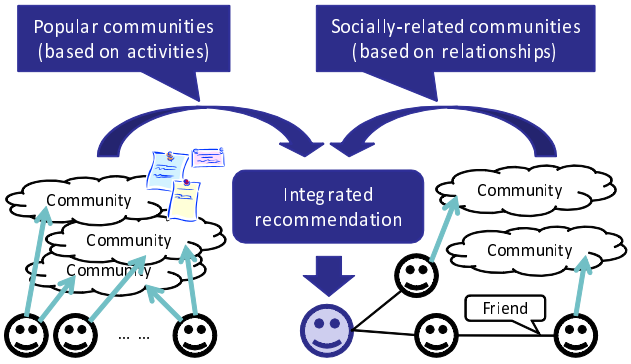

Figure 1: Overview of the methodology

3.2 Activities of Dissimilar Communities

As described in Section 2.3.1, popular communities are considered useful to be recommended. A popular community is the community that has many users or exchanges of information, so evaluating the number of members in the community or the number of messages posted in the community can be criterion for identifying them.

Some researches have proposed the methods to evaluate activity in the community [5, 6]. However, the correlation between the value of measurements and the users' interests is not clear. In addition, the reason of recommendation should be simple for the users, so the complicated methods are not appropriate especially for recommending dissimilar communities. Therefore, according to those researches, the following measurements are proposed to utilize for recommending dissimilar communities in this research.

- Number of members

  The number of members in a community indicates the number of users who are interested in the topic of the community. Thus, the community that has many users can be considered popular and versatile.

- Number of threads

  The number of threads in a community indicates the number of topics in it, so that the community which has many threads can be considered to be enough diverse to accept varieties of the users.

- Recent frequency of posting messages

  The recent frequency of posting messages in the community indicates the current activity of it.

The recent frequency, not the number of all the threads, is significant in the communities that are required to treat immediate or current affairs. In this research, the frequency is calculated by the duration of the latest 10 messages in the community.

3.3 Relationships with Dissimilar Communities

In the real world, almost all the users communicate with their acquaintances or friends, and obtain valuable or useful information through them. Consequently, many of them recognize human relationships are very important and who are the right people to get such important information. In SNSes, the users list their friends in the services, but those friends are not equivalent to the friends in the real world as described in 2.1. Thus, utilizing such relationships is possible and effective not only for recommending information, which the users' friends have, to the users, but also for deepening mutual understandings between them, which is not the scope of this research but is important for activating communications.

Meanwhile, a user can recognize only over the users that are at most two hops away from the user as socially-related people [7]. Thus, relationships of two hops on a social network should be utilized for recommending dissimilar communities. Accordingly, part of the communities a user's friends and friends of the friends join are recommended to the user in this research, as shown in Figure 1.

4 Experimental Results

4.1 Overview of the Experiment

For recommending dissimilar communities, the correlation between users' subjective estimations and the characteristics of communities, which are described in Section 3, should be unveiled to utilize two methods effectively. Thus, an experimental system has implemented for performing an experiment to investigate the correlation. This investigation requires real users and communities, so 20 examinees were collected from the users of mixi [8], which is one of the largest SNSes in Japan.

The investigation is performed as follows:

1. Random 10 communities are selected from all of the communities in mixi. If at least one of the examinee's friends or friends of friends joins in a community of the selected communities, the system selects communities again.
2. To investigate the relationships with dissimilar communities, random 10 communities are selected from the communities that at least one of the examinee's friends or friends of friends joins. Comparing the results between the random selected communities and these communities will enable us to identify the features of utilizing social relationships.
3. For evaluating whether the measurements of activities described in Section 3.2 are effective for selecting popular communities, the values of the three measurements are calculated for all of the selected 20 communities.
4. The system presents each examinee a questionnaire form. The form consists of the title of the selected communities and the following questions for each community. The list of the communities is permuted at random.
   - Interest in the community
     For evaluating the necessity of selected communities, the system asks the examinee to choose "I'm getting interested" or "I'm NOT getting interested" in the community.
   - Importance of the community
     Even if an examinee is getting interested in some of the communities, the importance of them is not the same. Thus, the system asks the examinee about the importance of the community by using 5-grade evaluation. The value of 5 indicates the community is most important for the examinee.
   - Unexpectedness of the community
     To activate communications, unexpected but interesting communities should be recommended to the users. Thus, the system asks the examinee about the unexpectedness of the community by using 5-grade evaluation. The value of 5 indicates the community is most unexpected for the examinee.

4.2 Evaluations of Activities of Communities

In order to study how the three measurements for estimating the activities of communities described in Section 3.2 can apply to recommend dissimilar communities, the correlation between the importance of the communities obtained from the experiment and the values of them is evaluated.

First, Figure 2 illustrates the correlation between the number of participated users in the communities and the amount of important communities. In this evaluation, the 4 or 5 value of the importance of a

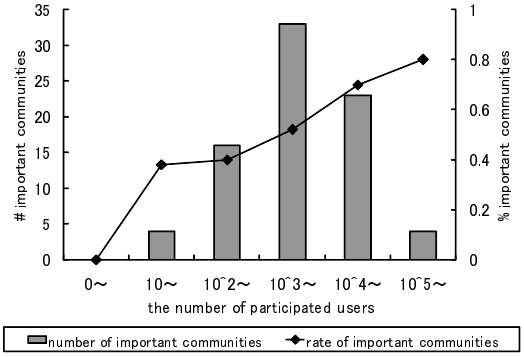

Figure 2: Correlation between the number of participated users and importance of communities

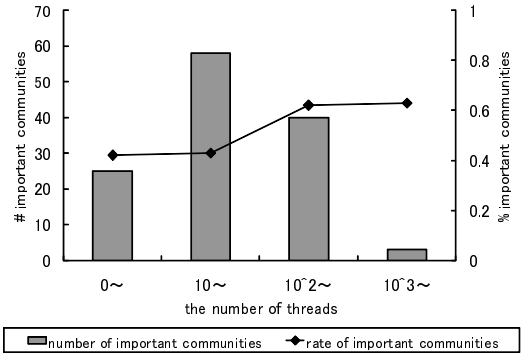

Figure 3: Correlation between the number of threads and importance of communities

community in the questionnaires, which means above average, is used for identifying the important community. As shown in Figure 2, the communities that have many users tend to be important. However, almost of the important communities do not have so many users.

Secondly, Figure 3 depicts the correlation between the number of threads in the communities and the amount of important communities. As shown in the figure, the number of threads does not affect the importance of the communities when the number of threads exceeds a specific threshold.

Finally, Figure 4 shows the correlation between the recent frequency of posting messages in the communities and the amount of important communities. As the frequency increases, the importance of the communities increases. However, almost of the important communities do not have so many users as well as the number of participated users.

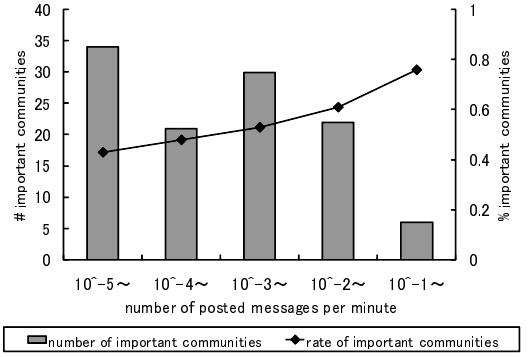

Figure 4: Correlation between the recent frequency of posting messages and importance of communities

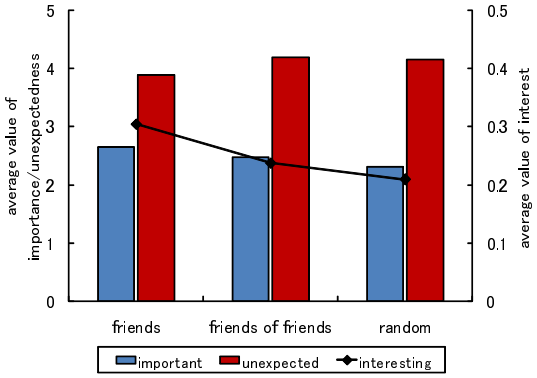

Figure 5: Differences of evaluations by utilizing social relations

4.3 Evaluations of Relationships with Communities

As described in Section 3.3, utilizing social relationships of the users is considered effective for recommending dissimilar communities. Therefore, in order to study the features of utilizing relationships, differences of the importance of the recommended communities between the communities selected at random and the communities selected based on the examinees' relationships are evaluated. The average values of the interest, importance, and unexpectedness of the communities are shown in Figure 5.

First, social distance of the examinees affects the interest of the communities. They tend to get interested in the communities their friends join, rather than the random communities. Secondly, the importance of the communities is the almost same in all the three cases. Lastly, the unexpectedness of the communities tends to high in all the three cases.

As a result, communities which are selected based on a user's social relationships tend to be unexpected,

i.e., dissimilar, and interesting communities.

4.4 Discussions

As described in Section 4.2, the measurements described in Section 3.2 can be considered to estimate the importance of communities. However, almost of the important communities do not have high values of the measurements. In addition, the communities which have high values of them tend to be the below the average value of unexpectedness. Thus, only utilizing the activities of communities is not enough to recommend dissimilar communities effectively.

Meanwhile, the evaluations of relationships described in Section 4.3 clarify the communities the users' friends join tend to be unexpected for the users as well as the communities selected at random. Although more than half of the recommend communities are useless for the users, the recommendation system can present the reason why these communities are recommended so that the users will be able to estimate the importance of communities easily based on the relationships with their friends and to deepen understandings of their friends. Thus, utilizing social relationships is effective both for recommending dissimilar communities and for activating communication environment in SNS.

Since these two methods have different effects, integrating both of them is expected to select appropriate dissimilar communities effectively. For example, when a part of the communities, which are selected based on the social relationships with the users, are extremely inactive, a recommendation system should exclude them. Therefore, the recommendation system that implements both of two methods is considered to realize fundamental activation of communications in SNSes.

5 Conclusion

In this paper, the necessity of activating communications in SNSes was described. Recommending dissimilar communities can be one of the methodologies to realize such activation. However, dissimilar communities tend to useless for the users, so determining appropriate dissimilar communities is important. Thus, the methodology that integrates the activities of communities and the relationships with the communities was proposed.

By carrying out the experiment, the characteristics of communities that are selected by the proposed methods were unveiled. As a result, the activities of communities are expected to select popular important communities, and the relationships with communities are expected to select unexpected dissimilar communities. Integrating both of two methods may enable the recommendation system to select appropriate dissimilar communities for the users.

However, the features of integration are not clear. Therefore, further investigation will be required to realize the integrated recommendation system.

Acknowledgements

This work was supported in part by the Research and Development Program of Ubiquitous Network Authentication and Agent (2003), The Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT), in Japan.

References:

[1] Google, "Orkut." available at `http://www.orkut.com/`.

[2] LinkedIn Corporation, "LinkedIn." available at `https://www.linkedin.com/`.

[3] J. Miller, D. Rosensweig, and E. Schmidt, "From player to platform: The context makes the connection," in *Transcipt of PC Forum 2004*, Mar. 2004.

[4] E. Spertus, M. Sahami, and O. Buyukkokten, "Evaluating similarity measures: A large-scale study in the orkut social network," in *Proc. of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 678–684, 2005.

[5] N. Matsumura, A. Miura, Y. Shibanai, Y. Ohsawa, and M. Ishizuka, "The dynamism of 2channel," *IPSJ Journal*, vol. 45, pp. 1053–1061, Mar. 2004. (in Japanese).

[6] N. Matsumura, "The dynamism of Yahoo! Japan message boards," in *Proc. of the 19th Annual Conference of the Japanese Society for Artificial Intelligence*, June 2005. (in Japanese).

[7] J. Goecks and E. Mynatt, "Leveraging social networks for information sharing," in *Proc. of the 2004 ACM conference on Computer supported cooperative work*, pp. 328–331, 2004.

[8] mixi, Inc., "mixi." http://mixi.jp/.