D-SNS: A Knowledge Exchange Mechanism Using Social Network Density among Mega-Community Users

Masao Ohira[1]
[1]Department of Information Science, Nara Institute of Science and Technology
8916-5 Ikoma, Nara, 630-0192, Japan
masao@is.naist.jp

Kumiyo Nakakoji[2, 3]
[2]RCAST, University of Tokyo
4-6-1 Komaba, Meguro, Tokyo, 153-8904, Japan
[3]SRA Key Technology Lab. Inc.
kumiyo@kid.rcast.u-tokyo.ac.jp

Ken-ichi Matsumoto[1]
[1]Department of Information Science, Nara Institute of Science and Technology
8916-5 Ikoma, Nara, 630-0192, Japan
matumoto@is.naist.jp

ABSTRACT
SourceForge.net (SF.net) is a large-scale online community hosting a number of open-source projects. We regard SF.net as a "*mega community*" because many of the members of SF.net are members of one or more open-source projects, which themselves form online communities. By regarding SF.net as a mega online community, we have identified four types of social relationships among the members of SF.net. This paper describes a mechanism to exploit measurement of the density of social networks to understand the nature of each type of social relationships in the mega community. We present D-SNS (Dynamic Social Networking System), which promotes knowledge exchange among SF.net users, by using measurement results of the density of social networks.

Categories and Subject Descriptors
H.5 [**Information Interfaces and Presentation**]: Group and Organization Interfaces – *collaborative computing, computer-supported cooperative work, organization design, web-based interaction*

General Terms
Management, Measurement, Human Factors

Keywords
Mega community, knowledge exchange system, social network analysis, density of social networks

1. INTRODUCTION
An online community such as open source software (OSS) community is a means to create artifacts by collaboration among community members. A number of studies have tried to understand various aspects of online communities [1][2][3]. The primary objectives of most of the studies are to reveal characteristics of collaboration and communication in a single community, and to construct methodologies and tools for supporting online communities. An increasing number of recent studies especially focus on social relationships among community members using social network analysis (SNA) in order to support activities of members in the community [4][5].

This paper presents our study that focuses on social relationships among people in a particular kind of online community, a "*mega community*." A mega community is a large-scale online community consisting of a number of smaller-scale communities. The interesting aspect of a mega community is where a member seems to naturally switch the role between a member of a mega community and a member of a smaller community. In fact, many of the members of a mega community we studied belong to more than one smaller scale communities [6]. As the result, there are different kinds of multiple social relationships in a mega community. The goal of our study is to support creation of social relationships suitable to various roles of members in a mega community.

In the next section, we illustrate SourceForge.net[1], which is an online OSS development community, as a representative example of a mega community. We then present the particularity of social relationships observable in SF.net. Section 3 describes a way to understand the nature of social relationships by measuring the density of social networks. Section 4 discusses how we can use measurement results of the density of social networks. We introduce the prototype system called D-SNS (Dynamic Social Networking System), that promotes knowledge exchange among SF.net users, as an application of exploiting measurement results of the density of social networks.

2. SourceForge.net: A MEGA COMMUNITY
SourceForge.net (SF.net) is a large-scale online community for OSS (Open Source Software) development. People register in SF.net to become a SF.net user. The SF.net users have a variety of roles (e.g., developers, bug reporters, end-users, donators and so on). The number of unique user accounts is over 1.2 millions in March 2006.

We regard SF.net as a "mega community," which is a community for a number of OSS communities. An OSS community in SF.net is called a "project". Over one hundred thousand OSS projects are registered on SF.net. SF.net users participate in activities of each OSS project such as releasing OSS, discussing issues on OSS development, reporting bugs and so forth.

The social relationships among SF.net users are created through the activities. SF.net users communicate with each other by using mechanisms such as bulletin board systems called forums, mailing lists, and bug reporting (tracking) systems. They rarely meet face-to-face. Here, social relationships are assumed to be

[1] SourceForge.net, http://sourceforge.net

relations emerged from results of communications among SF.net users.

There are discriminative social relationships in a mega community such as SF.net, which are different from social relationships in a single, common OSS community. Figure 1 simply illustrates four kinds of social relationships in SF.net as graphs called social networks. Graphs for social networks represent persons as nodes and relations between persons as lines (edges).

Figure 1-(a) depicts the social relationships among all SF.net users in case of viewing SF.net as a single community. The social relationships are defined as the same as that in a common OSS community.

Figure 1-(b) shows the social relationships in each project in SF.net. A small circle represents an OSS project as a single community. The social relationships are defined by relations created in each project.

Figure 1-(c) represents the social relationships which user(X) has in SF.net, in user(X)'s point of view. This type of social network is called ego-centric network. User(X) has connections to five users.

Figure 1-(d) shows the social relationships which user(X) has in each project in SF.net, in the user(X)'s point of view. Because user(X) participates in three projects, her/his social relationships are represented as three social (ego-centric) networks.

In this way, different types of social relationships exist in a mega community depending on ways of cutting off social relationships, that is, standpoints of people involved in the mega community (e.g., (a) is for administrators of a mega community, (b) is for managers of communities, and so on.). It would be important for people in a mega community to understand the nature of social relationships according to own roles or positions.

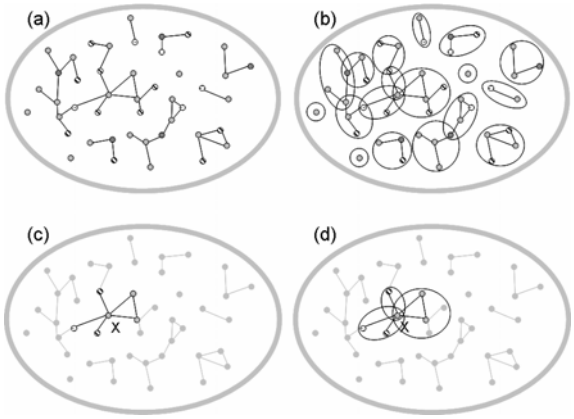

Figure 1. Four types of social relationships in a mega community

3. MEASURING SOCIAL RELATIONSHIPS

3.1 Characteristics of Social Relationships

There are two social networks extremely representing characteristics of social relationships. One is that called open network in the left of Figure 2. The other is closed network in the right of Figure 2.

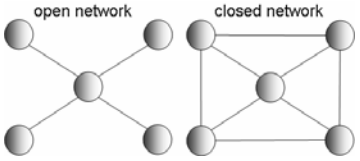

Figure 2. Open network and closed network

The characteristics of open network and closed network are described by [7] as follows.

> ***Open Network****: is a large, open, diverse, and externally focused network. It is excellent for getting lots of new information, learning about new opportunities, and finding resources. However, it is not so good for building consensus, producing consistent expectations, or developing a sense of common mission (may be prone to conflicts or tensions).*
>
> ***Closed network****: is a small, closed, homogeneous, and internally focused network. It is good for building group loyalty, identity, and a sense of common purpose. The disadvantage is that it may be inadequate for getting information or other resources, or insufficient for influencing people outside the networks. It is subject to group thinking and the development of an us-versus-them view of the world.*

3.2 Measuring Network Density

There are many metrics in the area of social network analysis [8][9], which can be used to know characteristics of social relationships in an organization or group. Measuring the density of social networks is a simple way to know whether social networks have the characteristics of open network or closed network [9]. If social networks with low density, the social networks tend to have the characteristics of open network. If social networks with high density, the social networks often have the characteristics of closed network.

The density of social networks is defined as the number of lines (edges) in social networks, expressed as a proportion of the maximum possible number of lines [8]. The formula for the density of social networks is

$$ND = \frac{2l}{n(n-1)}$$

where l is the number of lines (edges) in the networks and n is the number of nodes in the networks. The values of ND (network density) can be from 0 to 1.

As we described in Section 2, four types of social relationships exist in SF.net. Here, the network density can be defined for each type of social relationships.

(a) $ND(SF)$: ND in SourceForge.net
(b) $ND(P_i)$: ND in each project(P_i) in SF.net
(c) $ND(U_j, SF)$: ND of an ego-centric network each user(U_j) has in SF.net
(d) $ND(U_j, P_i)$: ND of an ego-centric network each user(U_j) has in each project(P_i) in SF.net

Table 1. Four types of network density in SF.net (Nov. 2005)

	ND	Max./Min.	σ^2
(a) *ND(SF)*	0.15 x 10^{-10}	N/A	N/A
(b) *ND(P_i)*	0.24 (avg.)	1.0/0	0.13
(c) *ND(U_j, SF)*	0.65 (avg.)	1.0/0	0.16
(d) *ND(U_j, P_i)*	0.66 (avg.)	1.0/0	0.17

No one ideal ND can fit all people in a mega community. Each ND is an indicator to understand the current state of social relationships in a community or around own, and to think how the social relationships ought to be in the future.

4. USING NETWORK DENSITY

The four types of *ND*s can be used to design support tools suitable to various roles or positions of people in a mega community.

For instance, for administrators of SF.net and managers of projects, *ND(SF)* and *D(P_i)* are respectively important clues to know the state of social relationships among users in SF.net and projects they have to manage. If a project manager thinks his project should be more closely united than the current, he would need tools for mediating or facilitating communications among his project's members, so that *ND(P_i)* will be higher than the current.

For each users in SF.net, measuring *ND(U_j, SF)* and *ND(U_j, P_i)* would be helpful to support them. If user(U_j) have social relationships expressing as high *ND(U_j, SF)*, she might want a help for finding people with whom she have never communicate before because she cannot get new information through her current social networks. A user who often has a conflict with other users in a project might hope *ND(U_j, P_i)* will be more higher.

Table 1 shows the result of analysis on four *ND*s in SF.net. The data for calculating four *NDs* is communication logs accumulated in forums (bulletin board systems). Using forums, SF.net users (e.g., developers, end-users, bug reporters, and so on) discuss issue related to OSS development. If user(U_A) posts a message in a forum for project(P_i) and user(U_B) replies the message, then it assumes that there is a social relation between user(U_A) and user(U_B) in project(P_i). We collected all messages in all accessible forums (1,230,000 communication logs among 160,000 SF.net users in 90,000 projects) and extracted social relationships among the users.

From the result in Table 1, *ND(SF)* is extremely low compared to other *NDs*. In case of viewing SF.net as a community, the social relationships among SF.net users are not close at all. In contrast to *ND(SF)*, the average of *ND(P_i)* is much higher (0.24). The result is natural because OSS development in SF.net proceeds through project-based activities and cross-project OSS development is infrequent [6]. Both *ND(U_j, SF)* and *ND(U_j, P_i)* show furthermore high values. This is because over 80% of all projects in SF.net consist of less than 3 developers [6].

The density of social network, which represents the nature of the social relationships in a mega community, varies according to where we are looking at in a mega community. Therefore, in order to support to build social relationships in a mega community, we would need to design tools in consideration of which aspects of social relationships we are trying to support.

5. D-SNS: AN APPLICATION

This section introduces the prototype system called D-SNS (Dynamic Social Networking System) that promotes knowledge exchange using social relationships among SF.net users, as an application of exploiting measurement results of the density of social networks. In this case, *ND(Uj, SF)* is considered to design the system. The detail of D-SNS is described in [6].

D-SNS collects communication logs in all accessible forums in SF.net and extracts information on social relationships among SF.net users, information on technical terms used for finding knowledgeable developers (i.e., information on who is knowledgeable about what), and information on communication frequency among users, from the communication logs.

D-SNS helps a user chose whom she should communicate with, according to the state of the user's social relationships. If a system's user inputs a question related to OSS into the system, the system finds other users knowledgeable on the question and recommends knowledgeable users who answer the question.

5.1 Mechanism

Figure 3 shows the mechanism of D-SNS. Here, suppose that user(U_A) with *ND(U_A, SF)*=0.7, who wants to make *ND(U_A, SF)* more lower, is asking a question. At first, user(U_A) inputs a question into D-SNS.

5.1.1 Searching knowledgeable people

D-SNS searches knowledgeable users using results of keyword matching between technical words stored in the system and keywords user(U_A) input. The system selects up to 20 users and delivers the questions to them.

5.1.2 Sorting by network density

If some users answer the question, D-SNS calculates *ND(U_A, SF)* after communicating with the users as Figure 4, and sorts the

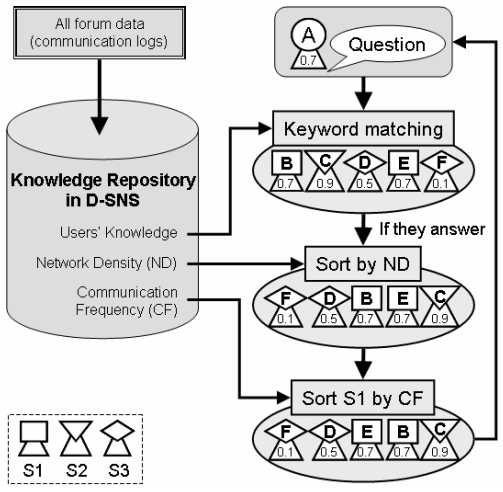

Figure 3. Mechanism of D-SNS

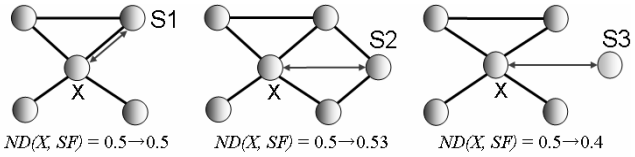

Figure 4. A communication partner and changes of network density

users in order of lower $ND(U_A, SF)$ in this case. Ordering depends on each user(U_j)'s preference.

5.1.3 Sorting by communication frequency

If $ND(U_A, SF)$ after communicating with knowledgeable users is same such as user(U_B) and user(U_E) in Figure 4, D-SNS sorts the users in order of higher communication frequency because user(U_E) who often communicates with user(U_A) might have better understandings of user(U_A)'s questions or demands than user(U_B).

5.2 User Interface

Figure 5 shows the user interface of D-SNS. The left of Figure 5 is for questioners (Alice). Alice can ask a question from the "Find People" tab. If someone replies the question, a list of respondents will appear as the list in the left of Figure 5.

The icons mean that S1 is a user within 1 degree of separation from Alice (i.e. have communicated with S1 before) as in the left of Figure 4, S2 is a user within 2 degrees from Alice as in the middle of Figure 4, and S3 is a user with more than 3 degrees of The numbers of the right of icons shows $ND(U_{Alice}, SF)$ if Alice communicates with the listed users.

In a similar way, a respondent (Ellen) can find questioners who would like to know Ellen's knowledge from the "Help People" tab in the right of Figure 5. If Ellen wants to tell a questioner (suppose Alice) something Ellen knows, Ellen can reply to Alice' questions using BBS only accessible for Ellen and Alice. If Ellen does not reply any questions, none of questioners can know that because questioners cannot know their questions will be delivered to whom.

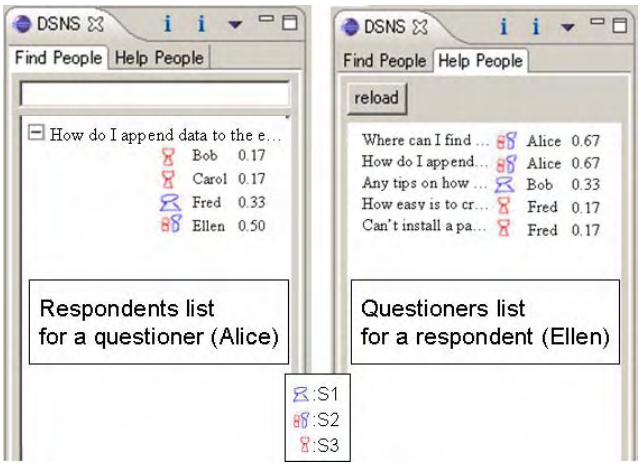

Figure 5. User interface of D-SNS

6. FUTURE WORK

In the near future, we need to enlarge data sources (e.g., mailing lists and bug reporting systems) for extracting social relationships and users' knowledge. We have a plan to elaborate ways of calculating *NDs* by using directed graphs or weighted graphs. We also would like to design support tools for administrators and managers in a mega community using $ND(U_j, SF)$ and $ND(U_j, P_i)$.

7. ACKNOWLEDGMENTS

This work is supported by the EASE (Empirical Approach to Software Engineering) community in the Comprehensive Development of e-Society Foundation Software program and Grant-in-aid for Scientific Research (B) 17300007, 2006 and for Young Scientists (B), 17700111, 2006, by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

8. REFERENCES

[1] Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P. and Kraut. R. E. Using social psychology to motivate contributions to online communities. In *Proceedings of the conference on Computer supported cooperative work (CSCW'04)*, 2004, 212-221.

[2] Millen, D. R. and Patterson, J. F. Stimulating social engagement in a community network. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'02)*, 2002, 306-313.

[3] Mockus, A., Fielding, R. and Herbsleb, J. D. Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 11(3), 2002, 309-346.

[4] Saltz, J. S., Hiltz, S. R., and Turoff, M. Student social graphs: visualizing a student's online social network. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'04)*, 2004, 596-599.

[5] Stefanone, M., Hancock, J., Gay, G., and Ingraffea, A. Emergent networks, locus of control, and the pursuit of social capital. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'04)*, 2004, 592-595.

[6] Ohira, M., Ohoka, T., Kakimoto, T., Ohsugi, N. and Matsumoto, K. Supporting Knowledge Collaboration Using Social Networks in a Large-Scale Online Community of Software Development Projects. In *Proceedings of APSEC2005 Workshop on Supporting Knowledge Collaboration in Software Development*, 2005, 835-840.

[7] Baker, W. E., Achieving Success through Social Capital. John Wiley & Sons Inc., 2000.

[8] Wasserman, S. and Faust, K. Social Network Analysis: Methods and Applications. Cambridge University Press, 1994

[9] Scott, J., Social Network Analysis: A Handbook. SAGE Publications, 2000.