Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2005 Proceedings

Americas Conference on Information Systems (AMCIS)

2005

Impact on Open Source Software Performance: A View from Social Structure

Yuan Long
CBA, ylong@unlserve.unl.edu

Keng Siau
University of Nebraska- Lincoln, siauk@mst.edu

Impact on Open Source Software Performance: A View from Social Structure

Yuan Long
University of Nebraska-Lincoln
ylong@unlserve.unl.edu

Keng Siau
University of Nebraska-Lincoln
ksiau@unlnotes.unl.edu

ABSTRACT

Open source software (OSS) has drawn increasing attention from both practitioners and researchers in recent years. However, few empirical studies have investigated the interaction process of OSS teams during OSS development. This research has two primary objectives. The first is to study the social structure of OSS teams. The second is to investigate the factors impacting the social structure of OSS teams as well as the effects of the social structure on OSS performance. Based on the social structure and social network theory, and group-related theories, this paper first presented a general research framework, then discussed research methodology, and finally describes a two-phase study plan. By exploring the relationships between group characteristics, social network, and OSS performance, the study contributes to the enhancement of knowledge on OSS development from the social structure perspective.

Keywords

Open Source Software (OSS), social network, social structure, social interaction, social network analysis, group theory

INTRODUCTION

The open source software (OSS) movement has attracted considerable interests from both practitioners and researchers in recent years. OSS covers a wide range of products – from operation systems (e.g., Linux), web browsers (e.g., Mozilla), email systems (e.g., Fetchmail), to programming languages (e.g., Perl). Besides products, OSS development process has also drawn considerable attention for its unique attributes of volunteerism, concurrent engineering, and rapid application development (Gallivan, 2001). However, few empirical studies have addressed the social aspects of OSS development process, especially the evolvement of social structure such as the interaction pattern and individual roles during the OSS development. To bridge this gap, this research in progress has two primary objectives. First, it aims to understand the social structure of OSS projects; second it purposes to explore the potential causes of the social structure (e.g., group composition, group size, and group task) as well as the impacts of the social structure on OSS performance.

The paper is organized as follows. First, a literature review is provided to introduce the current studies concerning the social structure of OSS projects. Second, group theory, social structure, and social network theory are discussed to serve as the theoretical foundation. After that, a general research framework is presented and the primary research methodology – social network analysis – is discussed. Finally, a two-phase study plan is proposed to achieve the research goals.

LITERATURE REVIEW

The phenomenon of OSS has caught increasing attention of both practitioners and researchers from diverse fields such as computer science, social psychology, organization, and management. Because of the complexity and multifaceted nature of OSS, researchers have investigated the OSS phenomenon from various perspectives. For example, focusing on the technical perspective, researchers studied issues such as OSS development methodology (e.g., Jørgensen, 2001) and coding quality (e.g., Stamelos, Angelis, Oikonomou and Bleris, 2002). Based on social psychology, researchers investigated individual motivation (e.g., Hann, Robert and Slaughter, 2004), social network (e.g., Madey, Freeh and Tynan, 2002), and social structure (e.g., Crowston and Howison, 2005). In terms of the organizational and managerial perspective, researchers examined knowledge innovation (e.g., Hemetsberger, 2004; Lee and Cole 2003; von Hippel and von Krogh 2003) and governance mechanism (e.g., Sagers, 2004).

OSS development team is essentially a virtual organization in which participants interact and collaborate with each other through the Internet. Comparing to conventional organizations, the structure of virtual organizations is suggested to be

decentralized, flat, and non-hierarchical (Ahuja and Carley, 1999). Raymond (1998) also stressed the decentralized nature of OSS development. However, some researchers challenge the belief that the social structure of OSS projects is decentralized and non-hierarchical in nature (e.g., Crowston and Howison, 2005; Gacek, 2004; Moon, 2000; Mockus, 2000, 2002). Instead, they argue that social structure of OSS projects is hierarchical rather than flat, either like a tree (Gacek, 2004) or like an onion (Crowston and Howison, 2005). The social structure directly influences the collaboration and decision-making process, and further affects the overall performance of OSS project as well as individual's perception of belongings and satisfaction. The debate is still on going and there is a paucity of research in this area. Therefore, one wonders what form of social structure may be present in OSS development, and what type of structure will emerge: centralized or decentralized, hierarchical or non-hierarchical, more onion-like or more tree-like, or a combination of the above depending on specific situation.

Focusing on the relationship between actors, social network analysis is a well-accepted approach to investigate interaction pattern. A few researchers have adopted this approach to examine OSS development. For example, Madey et al. (2002) examined the relationship between developers based on whether or not they belong to the same OSS project. However, the researchers did not study the relationship between developers in terms of their communication and collaboration (e.g., development activities such as bug reports and feature requests). Crowston and Howison (2005) examined the social structure of 120 projects from SourceForge.net, and concluded that OSS development teams vary widely in their centralization. Their research result suggests an inconsistence of group centralization across OSS projects. However, they only studied the interaction structure of each project at a certain point of time. To further understand the dynamics of social structure over time, longitudinal study is needed to provide a greater insight. In addition, the researchers only focused on one characteristic of social network—group centralization, while ignoring other equally important structure indices such as density and core/periphery fitness, which are critical to describe the shape of the social network. Instead of a static perspective, longitudinal data will be employed in this study to examine the evolvement of social structure of OSS projects from a dynamic perspective.

To summarize, this research addresses the following questions.

1. What are the social structures of OSS teams during OSS development? Specifically, how does the interaction pattern of OSS teams evolve over time?
2. What factors impact the formation of social structure for OSS projects? For example, is there any difference on social structure between projects with different size, group composition, and software type?
3. How does the social structure impact OSS performance?

THEORETICAL FOUNDATION

Group-related Theories

Germain (1999) provided an outline to study three basic aspects of groups: group structure, dynamics, and development.

Group structure refers to the relatively stable patterns of "interpersonal behaviors and expectations about behaviors constitute the group structure" (Germain, 1999). Germain (1999) proposed four major structural features of groups: social roles, norms, subgroups, and size and composition. Social roles refer to the defined positions with specific rights and duties. Norms refer to the shared values and rules. When individuals commit to a set of shared values, norms function as rules to regulate the behavior of members. Researchers also observed that size and composition (such as age and gender) influence group performance.

Second is group dynamics. Germain (1999) addressed the action of group members, interactions among group members themselves, and between group members and their environment. Group dynamics involve several major group processes such as leadership, communication, decision-making, cohesion, and morale. Cohesion here refers to the strength attracting members to the group, which functions as glue to hold the members together.

The last aspect is group development, which focuses on continuous group change. Group evolves with the emergence of new group members, goals, and even norms.

OSS projects are developed by a group of volunteers who communicate and collaborate with each other through the Internet. Drawing on group theory, our research objectives focus on dynamics and development of social network, specifically the interaction pattern of OSS teams during OSS development. In addition, our study will investigate the effects of group characteristics (e.g., group size and composition) on interaction pattern.

Social structure and social network

Social structure, as Schaefer and Lamm (1998) suggested, refers to the way in which society is organized into predictable relationships. Social structure can be considered in terms of three aspects— actors, actors' actions, and actors' interactions. Social actor is a relatively static concept addressing issues such as roles, positions, and status. Individual actor is embedded in the social environment and therefore his/her actions are largely influenced by interaction with other actors. Social interaction can be generally regarded as the way in which people respond to one another. These interaction patterns are to some extent independent of individuals. They exert a force that shapes both behavior (i.e., actions) and identity (i.e., actors) (Schaefer and Lamm, 1998).

Researchers of social interaction have been focusing on how individuals actually communicate with each other in group settings. The studies address issues such as interaction pattern, underlying rules guiding interaction, reasons accounting for the way people interact, and the impact of interaction pattern on individual behavior and group performance. These studies begin by questioning what might be the interaction pattern in a specific social setting. And that addresses our first research question — understanding social interaction of OSS projects.

Social network theory is based on the intuitive notion that social interaction patterns are essential to the individuals who reflect them. Network theorists believe that how individuals behave largely depends on how they interact with each other and how they are tied to the social network. Furthermore, theorists also believe that the success or failure of societies and organizations often depends on the internal interaction pattern (Freeman, 2002).

Besides the theoretical essence, social network theory is also characterized by a distinctive methodology encompassing techniques for data collection, statistical analysis, and visual representation. This approach is called social network analysis and will be discussed in the research methodology section.

RESEARCH FRAMEWORK

This research draws on social network theory and group theory to study the social network, its causes, and its impact on OSS performance.

Figure 1 shows the general framework of our research. The framework argues that social network, as an intermediate variable, is both an effect of group characteristics and a cause of OSS performance.

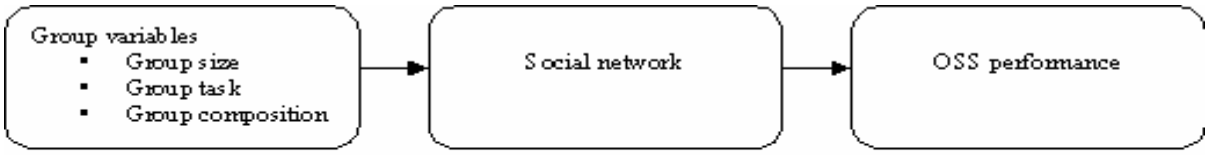

Figure 1. A Preliminary Research Framework

RESEARCH METHODOLOGY

Social network analysis

The research methodology employed in this study is social network analysis, with an aim of understanding the interaction pattern of OSS development process.

Guided by formal theory organized in mathematical terms and grounded in the systematic analysis of empirical data, social network analysis focuses on uncovering the interaction pattern of interdependent individuals (Freeman, 2004). Through a structural analysis of a social network diagram—a map depicting actors as well as ties between them—social network analysis offers promise to reveal the patterns of relationships and the relative positions of individuals in a specific social setting. This approach has been effectively used in organizational research, social support, and mental health (Freeman, 2004). It is adopted here for two primary reasons.

First, the focus of social network analysis matches our research objective. Social network analysis is focused on the relationship among a set of actors instead of their attributes. Our research aims to reveal the interaction pattern of OSS projects. Therefore, this approach offers great potential to answer our research questions.

Second, the availability of huge and rich data source on OSS projects provides sufficient and convenient resource to conduct social network analysis. Social network analysis is grounded in empirical data. Most OSS projects have online mailing lists, forums, and tracking systems that are open to the public. Thus they provide a rich set of longitudinal data source. Based on these public data sets, researchers are able to generate input data sets (i.e., matrix depicting relationships between each pair of actors) for social network analysis.

Data collection

Data will be collected from websites that host open source software projects. Among all these websites, SourceForge is the most popular one and is regarded as the world's largest open source software development website. SourceForge provides free tools and services to facilitate OSS development. Currently it hosts a total of 99,730 OSS projects and involves 1,066,589 registered users (data retrieved on May 4th, 2005 from the SourceForge.net). Data will be extracted from open forums such as bug-tracking system and feature requests as well as mailing list and other archives such as Concurrency Version Systems (i.e., a place to manage the codes).

Measurement

Group variables, such as group size, will be retrieved directly from the SourceForge (or other websites hosting OSS projects). Group task can be simply induced from the project type. Group composition can be gathered from the archives of the project.

Social interaction data can be extracted from archives of OSS projects. For example, messages from the bug-tracking forum will be coded to study the interaction among participants during bug-fixing procedure. A bug-tracking forum is an online discussion board containing bug reports and problem resolutions. Each message posted in that forum has a sender and a receiver. The reply of a message could be viewed as a link between the sender and the receiver. After coding each of these messages, a matrix will be generated to depict the interaction between each pair of participants. The matrix will be used as the input data for social network analysis. Social network analysis software will then be used to analyze network variables, such as group centralization and density.

OSS performance can be measured using instruments developed in earlier literature (e.g., success measurement discussed by Crowston, Annabi, and Howison, 2003).

RESAEARCH DESIGN

A two-phase study is planned for this research. The first stage is to understand the social structure of OSS projects. A social network analysis approach will be employed to study the interaction pattern during OSS development process. Interaction data will be extracted from websites hosting OSS projects. Critical network variables such as group centralization, density, and cliques will be calculated by social network analysis software. Specific hypotheses will be generated based on an in-depth understanding of social structure of OSS teams at the end of this stage. The first stage will be completed before August 2005.

In the second stage, a full-scale research study will be conducted to test the hypotheses proposed in stage I. Questionnaires and interviews will be used in this research. Using both questionnaires and interviews will provide the necessary triangulation. Statistical techniques will be employed to analyze the relationships among variables. This step is expected to be completed by the middle of 2006.

REFERENCES

1. Ahuja, M. and Carley, K. (1999) Network Structure in Virtual Organizations, *Organization Science*, 10(6), 741-747
2. Crowston, K., Annabi, H., and Howison, J. (2003) Defining Open Source Software Project Success. Proceedings of ICIS 2003, Seattle, WA, 14-17
3. Crowston, K. and Howison, J. (2005) The social structure of Free and Open Source software development, *First Monday*, 10(2)

4. Freeman, L.C. (2002) The study of Social networks. http://www.insna.org/INSNA/na_inf.html
5. Freeman, L.C. (2004) *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver: Empirical Press.
6. Gacek, C., & Arief, B. (2004) The many meanings of open source. *IEEE Software, 21*(1), 34–40.
7. Gallivan, M.J. (2001) Striking a balance between trust and control in a virtual organization: a content analysis of open source software case studies. *Information systems journal*, 11, 277-304.
8. Germain, C.B. & Bloom, M. (1999) *Human behavior in the social environment: An ecological view* (2nd ed) New York: Columbia University Press.
9. Hann, H., Robert, J. and Slaughter, S. (2004) Why Developers Participate in Open Source Software Projects: An Empirical Investigation. In *Twenty-Fifth International Conference on Information Systems*, Washington, DC, 821-830.
10. Hemetsberger, A. (2004) Sharing and Creating Knowledge in Open-Source Communities: The case of KDE The Fifth European Conference on Organizational Knowledge, Learning, and Capabilities in Innsbruck, Austria.
11. Jørgensen, N. (2001) Putting it all in a trunk: incremental software development in the FreeBSD open source project. *Information Systems Journal*. 11, 321-336.
12. Lee, G.K. and Cole R.E. (2003) From a Firm-Based to a Community-Based Model of Knowledge Creation: The case of the Linux Kernel Development. *Organization Science*, 14, 6, 633-649.
13. Madey, G., Freeh, V., and Tynan R. (2002) The Open Source Software Development Phenomenon: An Analysis Based on Social Network Theory, AMCIS2002, Dallas, TX.
14. Mockus, A., Fielding, R. T., and Herbsleb, J. D. (2000) A case study of open source software development: The apache server. Proceedings of ICSE 2000
15. Mockus, A., Fielding, R. T., and Herbsleb, J. D. (2002) Two case studies of open source software development: Apache and mozilla. ACM Transactions on Software Engineering and Methodology, 11, 3, 309–346.
16. Moon, J. Y., and Sproull, L. (2000) Essence of distributed work: The case of linux kernel. *First Monday*, 5, 11.
17. Raymond, E.S. (1998) *The Cathedral and the Bazaar*. http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/. Accessed in Jan. 2005
18. Sagers, G.W. (2004) The influence of network governance factors on success in open source software development projects. In *Twenty-Fifth International Conference on Information Systems*, Washington, DC, 427–438
19. Schaefer, R. T. and Lamm, R.P. (1998) *Sociology*. 6th edition. McGraw-Hill.
20. Stamelos, I., Angelis, L., Oikonomou, A., and Bleris, G.L. (2002) Code quality analysis in Open-Source software development. *Information Systems Journal*, 12, 1, 43-60.
21. Von Hippel, E., G. and Von Krogh. (2003) Open source software and the "Private-Collective" Innovation Model: Issues for Organization Science. *Organization Science*. 14, 209-223.