December 2006

Impact of Collaborative Ties on Open Source Software Development

Chen Zhang
Krannert Graduate School of Management Purdue University

Impact of Collaborative Ties on Open Source Software Development

Chen Zhang
Krannert Graduate School of Management
Purdue University
zhang153@purdue.edu

ABSTRACT

Within the past few years, the open source software development (OSSD) has gained tremendous popularity as an alternative and promising approach to software production. The main objective of this study is to examine the impact of collaborative relations on the behavior of the OSSD team. In this proposal we develop a theoretical framework and research hypotheses. We then outline the research methodology of empirically testing the hypotheses on a large-scale data set. The research will be conducted in two stages. In the first stage, we will examine the impact of collaborative ties on developer joining behavior and team formation process. In the second stage, we will investigate the relationship between the resultant structural properties of the team and project performance.

Keywords

Collaborative network; open source software (OSS); open source software development (OSSD); social capital; structural holes; network closure.

INTRODUCTION

The creation of industrial-strength software has traditionally been conducted and managed within a firm setting. In recent years, however, an alternative model of software development, the open source software development (OSSD) in which programmers in Internet-based communities voluntarily collaborate to produce software, has emerged as a promising approach to developing high-quality software (Raymond 2001).

Despite the impressive success of some OSSD projects, many projects fail to take off and become abandoned. One of the main reasons for the failure of these projects may be the lack of developers in the project team. Since typically OSSD projects do not offer monetary rewards for developers to join and contribute, many projects are under-staffed and consequently are not well-equipped to deal with the complexity in software development. Hence it is important to understand how developers self-organize into project teams. Furthermore, although it has been recognized that OSSD has become a significant social phenomenon and that open source software (OSS) developers and users form a complex social network via various electronic communication channels over the Internet (von Hippel and von Krogh 2003), only recently have researchers begun to examine this phenomenon from a social network perspective (Lopez-Fernandez, Robles, and Gonzalez-Barahona, 2004; Madey, Freeh, and Tynan, 2002; Xu, Gao, Christley, and Madey, 2005). However, these works tend to focus on the technical properties of the OSSD network. Therefore, in this dissertation we focus on the dynamic process of project team formation within the OSS collaborative network and the performance implications of the resultant structural patterns of the team.

THEORETICAL FRAMEWORK

We will conduct two inter-related studies. In the first study, we examine the impact of social ties on project team formation. In the second study, we investigate the relationship between structural properties of the OSS teams and project performance. Our research model is presented in Figure 1.

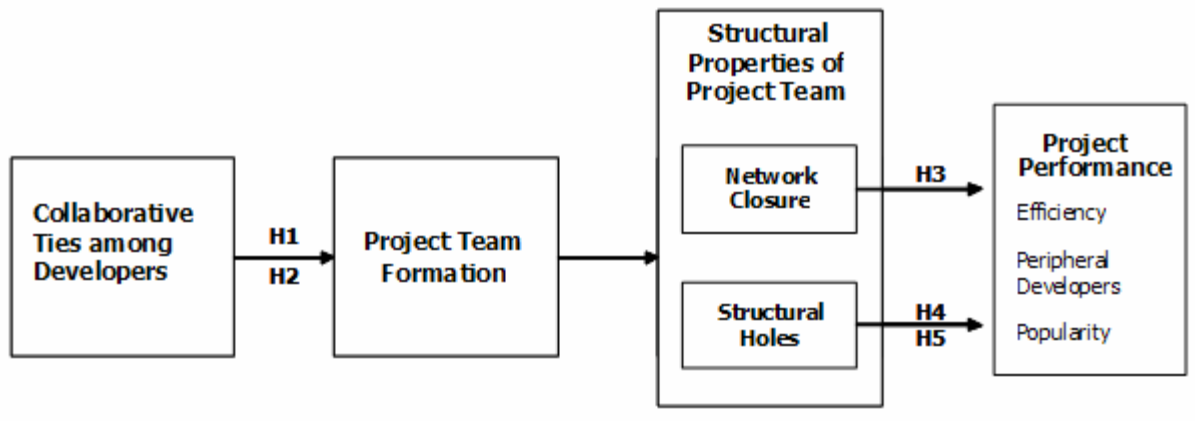

Figure 1. Research Model

Study 1 – Impact of Collaborative Ties on Open Source Software Development Team Formation

Just as the social position of a firm within inter-organizational networks influences its alliance strategies and consequent outcomes (Gulati, 1995; Powell, Koput, and Smith-Doerr, 1996), we argue that social relations forged during past collaborations will have an impact on how OSS project teams take form. In this study, we empirically examined the role of collaborative ties in OSSD team formation and developer joining behavior.

Prior literature suggests people are more likely to work together when there exist social ties between them (Schachter 1959). Furthermore, teams consisting of individuals with preexisting relationship ties have been shown to solve complex problems better than teams of strangers because they are able to pool information more efficiently (Gruenfeld, Mannix, Williams, and Neale, 1996). In the open source context, given the lack of opportunities of face-to-face contact, developers would be quite concerned about effective communication and coordination. Hence, if a developer has past collaborative relations with the existing members of a project, he or she can be assured that coordination and communication with other team members would be more efficient and effective due to the shared understanding accrued from prior interactions. We propose the following hypotheses with regard to the impact of preexisting collaborative ties on the OSSD team formation:

H1*: From a project standpoint, projects whose initiators have social ties with other members in the OSSD network are more likely to have other developers join subsequently than projects whose initiators do not have social ties.*

H2: *From a developer standpoint, whether a developer has strong ties with a project's initiators positively influences the probability of the developer joining the project.*

Study 2 – Impact of Structural Properties on the Performance of Open Source Software Development Teams

After examining the impact of collaborative ties on developers' decisions to join an OSS project, we further investigate whether or not the resultant network structure of the project team has a significant effect on project performance. In order to answer this question, we identify two relevant structural properties – network closure and structural holes - that have been frequently studied in the sociology literature. We propose that network closure or cohesiveness of the project team contributes to easier coordination and more effective communication, leading to higher production efficiency. A loosely-connected network rich in structural holes enables the project to have access to a larger developer base or user pool that would be otherwise unreachable. Consequently, the project is more likely to receive code contributions from developers outside the project team. Additionally, it tends to be more popular in terms of number of downloads and visits.

According to social network theorists, network closure or densely connectedness of networks can facilitate the production of social norms and promote trust (Coleman, 1988, 1990). Since social capital depends on the cooperation among actors, which is typically impaired by lack of trust, a closely-knit dense network tends to have more social capital due to higher level of trust and cooperation. In the open source context where a group of geographically-dispersed developers volunteer to jointly produce software, the production process relies even more heavily on the successful collaboration and coordination among developers. A densely connected project team is likely to incur less conflicts and misunderstandings between members and to communicate more effectively, which then leads to more efficient production process. Therefore we develop the next research hypothesis:

H3: *Cohesiveness of an OSS team is positively related to the performance of the project in terms of production efficiency.*

On the other hand, structural hole theory (Burt 1992) stresses that social capital is derived from the diversity of information and the brokerage opportunities across separate loosely-connected social clusters. Networks rich in structural holes enable some members to have access to many distinct information flows and a larger resource pool. In the open source network if a project team consists of members who have ties with some remote social clusters, it tends to attract more developer and user attention than a team spanning fewer distinct social clusters. Consequently the project is likely to receive code contributions or suggestions from other programmers, whom we define as peripheral developers. And it tends to be downloaded and visited more often. Based on this line of arguments we propose the following:

H4: *Number of structural holes in an OSS team is positively related to the number of peripheral developers who contribute code or ideas to the project.*

H5: *Number of structural holes in an OSS team is positively related to the popularity of the project in terms of number of downloads and the number of visits to the project site.*

RESEARCH METHODOLOGY

Data were collected from the open source software projects hosted on SourceForge.net. We randomly selected 1030 new projects that were registered between September and November in 2005. A web crawler downloaded the HTML files containing project summary data and developer information on the date of registration. We revisited sample projects one month after their registration to identify those developers who have subsequently joined. This process enables us to distinguish between the initiator and the developers who subsequently join. Further, in order to identify the social ties of the developers, we further collected data on the other projects that each developer has participated in to identify his or her past collaborators. Based on this data, we are able to construct affiliation matrices of developers and projects that depict the existence of the relationship ties between developers.

When testing hypotheses H1 and H2, we conducted the analyses at two levels – project level and developer level. At the project level, we controlled for the number of initiators, experience of initiators, and the ambiguity level of the project definition. At the developer level, the unit of analysis is a project developer dyad. We constructed the sample using choice-based sampling because of rare event data. Further, we controlled for the technical fit between project requirements and developer skills because technical similarity (e.g. operating system, programming language, and topic) between a project and a developer's background may strongly affect the probability of the developer joining the project. We also took into consideration the number of project initiators, experience of initiators, and experience of the developer.

In order to test hypotheses H3, H4, and H5, we need to collect project performance and activity data as well as the entire set of network data from SourceForge.net in order to compute the structural properties of sample projects.

RESULTS TO DATE AND PLANS FOR COMPLETION

Our results show support for hypotheses H1 and H2. At the project level, we found that the existence of prior social relations in the network does increase the probability of an OSS project to attract more developers. At the developer level, we found that a developer is more likely to participate in a project whose initiators have collaborated with him or her in the past.

We are currently in the process of gathering data from SourceForge.net and its monthly dumps of databases to the University of Notre Dame. Analyses of the performance implications of project team structural properties are expected to be completed by November 2006.

CONTRIBUTIONS

Overall this dissertation will advance the understanding of the OSS movement as a social phenomenon by identifying the role of collaborative relations. It has several contributions. First, we fill a gap in the open source literature by conducting an investigation of the role of social ties on project team formation patterns. Second, the adoption of social network analysis, which has received little attention in the OSS literature, can yield some interesting results on the interactions among OSS developers. Third, this study will shed some light on the dynamic socialization process in OSS and offer some insights to IT managers who intend to adopt open source as an organizational form. Overall, we expect this study to have theoretical, methodological, and practical implications.

REFERENCES

1. Burt, R. (1992) Structural holes: The social structure of competition. Harvard University Press, Cambridge.
2. Coleman, J.S. (1988) Social capital in the creation of human capital, *American Journal of Sociology*, 94, S95-S120.
3. Coleman, J.S. (1990) Foundations of social theory. Harvard University Press, Cambridge, MA.

4. Gruenfeld, D.H., Mannix, E.A., Williams, K.Y., and Neale, M.A. (1996) Group composition and decision making: How member familiarity and information distribution affect process and performance, *Organizational Behavior & Human Decision Processes*, 67, 1, 1-15.
5. Gulati, R. (1995) Social structure and alliance formation pattern: A longitudinal analysis, *Administrative Science Quarterly*, 40, 619-652.
6. Lopez-Fernandez, L., Robles, G., and Gonzalez-Barahona, J.M. (2004) Applying social network analysis to the information in CVS repositories, in Ahmed E. Hassan, Richard C. Holt, and Audris Mockus (Eds.) *Proceedings of the First International Workshop on Mining Software Repositories (MSR 2004)*, May 24, Edinburgh, Scotland, UK.
7. Madey, G., Freeh, V., and Tynan, R. (2002) The open source software development phenomenon: An analysis based on social network theory, in *Proceedings of the Eighth Americas Conference on Information Systems*, August 9-11, Dallas, Texas, USA.
8. Powell, W.W., Koput, K.W., and Smith-Doerr, L. (1996) Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology, *Administrative Science Quarterly*, 41,116-145.
9. Raymond, E.S. (2001) The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary. O'Reilly & Associates, Sebastapol, CA.
10. Schachter, S. (1959) The psychology of affiliation. Stanford University Press, Stanford, California.
11. von Hippel, E., and von Krogh, G. (2003) Open source software and the "private-collective" innovation model: Issues for organization science, *Organization Science*, 14, 2,209-223.
12. Xu, J., Gao, Y., Christley, S., and Madey, G. (2005) A topological analysis of the open source software development community, in *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS 05)*, January 3-6, Hawaii, USA.