Impact of Social Ties on Open Source Project Team Formation

Jungpil Hahn[1], Jae Yoon Moon[2], and Chen Zhang[1]
1 Krannert School of Management, Purdue University
West Lafayette, IN 47907, USA
{jphahn, zhang153}@mgmt.purdue.edu
2 Business School, Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
jmoon@ust.hk

Abstract. In this paper, we empirically examined the role of social ties in OSSD team formation and developer joining behavior. We find that the existence and the amount of prior social relations in the network do increase the probability of an OSS project to attract more developers. Interestingly, for projects without preexisting social ties, developers tend to join the project initiated by people with less OSSD experience. This research fills a gap in the open source literature by conducting an empirical investigation of the role of social relations on project team formation behavior. Furthermore, 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.

1 Background and Motivation

The creation of industrial-strength software code (or software development) has traditionally been regarded as an activity that can only be effectively conducted and managed within a firm setting. Recently however, an alternative model of software development, the open source software development (OSSD) in which programmers in Internet-based communities collaborate to voluntarily contribute programming code, has emerged as a promising approach to developing high-quality software [1]. During the past few years, a number of open source software (OSS) products, ranging from end-user applications (e.g., Emacs and OpenOffice), programming languages (e.g., Perl and PHP) to applications supporting the Internet infrastructure (e.g., sendmail), have been widely adopted. The prominence garnered by well-known OSS projects such as the Apache Web Server and the Linux operating system kernel are testimonies to the attractiveness and viability of OSSD as an alternative to the conventional proprietary model of producing software [1-3].

Despite the impressive success of some OSSD projects, it is a harsh reality that the vast majority of OSS projects fail to take off and become abandoned. One of the main reasons cited for the failure of OSS projects is the lack of developers in the project teams, or the inability of the project to bring together a critical mass of developers [2, 4]. Since it is typically the case that OSSD projects do not provide monetary rewards for developers' contributions, many OSSD projects are under-

Please use the following format when citing this chapter:
Hahn, J., Moon, J.Y., and Zhang, C., 2006, in IFIP International Federation for Information Processing, Volume 203, Open Source Systems, eds. Damiani, E., Fitzgerald, B., Scacchi, W., Scotto, M., Succi, G., (Boston: Springer), pp. 307-317

staffed and consequently are not well-equipped to deal with the complexity in software development. Hence, in order to understand and solve the key problems related to staffing, it is important to understand the dynamics of software team formation – how developers self-organize into project teams.

In this paper, we undertake an empirical examination of the formation of OSS project teams from a social network perspective. The OSSD community is essentially a complex collaborative social network endowed with social capital. Just as the social position of a firm within inter-organizational networks influences its alliance strategies and consequent outcomes [5-6], we argue that social relations forged during past collaborations will have an impact on how OSS project teams take form. However, despite the apparent relevance and importance of social capital in OSSD, only a few studies have examined its impact on developer behavior in team formation from a social network perspective. In this paper, we ask ourselves whether the existence and amount of prior social ties in an OSS project helps it attract additional developers. The remainder of this paper is organized as follows. In section 2, we present our theoretical background and develop our research hypotheses. We outline the empirical research methodology in section 3 and present the results in section 4. We conclude in section 5 by discussing the implications, contributions and directions for future research.

2 Theoretical Background and Research Hypotheses

This study draws from two streams of research – 1) open source software development (OSSD), and 2) social network analysis and network structure. We review and synthesize the relevant literature to develop our research hypotheses.

2.1 Open Source Software Development (OSSD)

Since its emergence, OSSD has posed many interesting questions for researchers in many fields. A number of researchers have addressed the factors that motivate individuals to participate in OSSD despite the lack of monetary compensation. Among the possible explanations for developers' participation in OSS projects are incentives related to career concerns and ego gratification [7]. Hars and Qu [8] identify both intrinsic motivations such as altruism and extrinsic motivations such as direct compensation. Another study surveys the motivations of the contributors to a large OSS project and finds that participation is mainly driven by developers' group identification, by the possibility of improving their own software, and by their tolerance of the required time investments for contributing to the project [9]. Lakhani and Wolf [10] identify enjoyment-based intrinsic motivation, user need, and learning as the most pervasive drivers of developer participation. In summary, the studies suggest that developers participate in OSSD mainly because of intrinsic factors such as enjoyment and extrinsic factors such as career advancement. However, the motivations identified from these surveys of developers do not explain why

developers choose to join one project over other possible similar projects. When deciding whether to join an OSS project, in addition to the previously cited motivational factors, a developer will also be concerned about issues related to coordination and communication with other team members. In general, when forming teams people prefer to work with those with whom they have worked in the past [11]. Familiarity bred from preexisting social relations with others can facilitate the newcomer's socialization process. Hence, we identify and test social ties among developers as a potential driver behind developer joining behavior and project team formation.

2.2 Social Network Analysis

Social network analysis aims to understand the relationships between people, groups, organizations, and other types of social entities [12-14], and has been used extensively in fields such as sociology [13, 15] and management [16-17] among others [18-19]. A social network is modeled as a graph with nodes representing the individual actors in the network and ties representing the relationships between the actors.

In a social network the actors maintain a tie by exchanging either tangible or intangible resources such as information, goods and services, and financial support. The strength of a social tie varies depending on a number of factors. Granovetter [12] distinguishes between strong and weak ties and asserts that tie strength depends on the amount of time, the emotional intensity, the intimacy, and the reciprocal services associated with the relationship. Strong ties are characterized by a sense of special relationship, an interest in frequent interactions, and a sense of mutuality of the relationship [20]. In contrast, weak ties are maintained infrequently or indirectly between the actors who belong to different social clusters. Both strong ties and weak ties play an important and differential role in a social network. Strong ties maintain and promote trust and collaboration whereas weak ties enable actors to access resources and information that are unavailable in their immediate social circles [12, 21].

2.3 Social Network Perspectives of Open Source Software Development

Although it has been recognized early on that OSSD has become a significant social phenomenon and that OSS developers and users form a complex social network via various electronic communication channels on the Internet [22], few researchers have examined this phenomenon from a social network perspective. Madey, Freeh, and Tynan [23] conducted the first empirical investigation of the open source movement by modeling OSS projects as a collaborative social network and found that the OSSD community can be modeled as a self-organizing social network. Others propose the methodology of applying social network analysis to data gathered from CVS code repositories of OSS projects [24]. Xu, Gao, Christley, and Madey [25] explored some social network properties in the open source community to identify patterns of

collaborations. However, the works cited above tend to be highly technical and mainly investigate the network properties of the OSSD community, offering limited theoretical and practical contributions. The work most similar to our research is done by Ducheneaut [26] who examined the socialization process of newcomers over time as a learning process and a political process by analyzing the developer activities in a large OSS project.

2.4 Research Hypotheses

Conventionally, project teams are formed by a manager assigning individuals to a team based on certain characteristics such as expertise and personality. An alternative approach is driven by team members' self-selection into teams. As such, in OSSD, some project initiators may formally recruit developers[1] (e.g., by broadcasting position openings and required qualifications to the entire community), or alternatively developers may voluntarily join a project team or be invited to participate in a project team by its existing members. Prior research suggests that people are more likely to work together when they have prior social ties [27-28]. Moreover, teams consisting of individuals with preexisting relationships have been shown to solve complex problems better than teams of strangers because they are able to pool information more efficiently [29]. In the open source software development context in particular, due to the lack of opportunities for face-to-face contact, developers face greater barriers to effective communication and coordination and are thus more likely to be concerned about these issues. Direct social relations with existing members of a project can mitigate concerns regarding communication and coordination difficulties due to the shared context accrued from prior interactions. We propose the following hypothesis with regard to the impact of preexisting social ties on open source software development project team formation:

H1: *Projects whose initiators have preexisting social ties with the network are more likely to have other developers join the development team than those whose initiators do not have ties.*

Projects can fall into two categories depending on whether or not their initiators have relationship ties in the network. Some projects are initiated by developers who have participated in other projects and formed social relationships with other developers in the community. For this type of project, the more social ties the initiators have, the larger will be the pool of potential developers. Consequently, these projects will be able to attract or invite others into the development team more easily.

[1] Interestingly, the extent of recruiting is surprisingly low based on our informal observations. For example, there are only about 200 position openings posted on SourceForge.net. When we consider that there are currently over 100,000 OSS projects are hosted on SourceForge.net, this number is quite inconsequential.

Therefore, we propose the following hypothesis regarding the impact of the amount of preexisting strong ties in a project:

H2: For those projects whose initiators have preexisting social ties with the network, the amount of such ties is positively associated with the probability of having other developers join the project team.

It may not necessarily be the case that projects are initiated by developers who are well connected to the network. Some projects may be initiated by developers who have yet to collaborate with others in the open source software development community even though they may have experience in managing software projects before (i.e., self-developed projects). In such cases, developers with prior open source project experience will have superior knowledge of OSS development and management processes. As a result, projects initiated by developers with prior experience may be more likely to have additional team members than projects initiated by developers with no prior experience. We propose the following hypothesis:

H3: For those projects whose initiators do not have preexisting social ties in the network, the experience of initiators is positively associated with the probability of having other developers join the project team.

3 Results

3.1 Data Collection and Measures

We collected data from open source software projects hosted on SourceForge.net. As the largest repository of open source applications on the Internet, SourceForge.net currently provides free hosting to more than 100,000 projects and more than 1,100,000 subscribers. It also offers a variety of services to hosted projects, including site hosting, mailing lists, bug tracking, message boards, file archiving, and other project management tools. SourceForge.net has been an attractive source of data for many researchers studying open source software mainly due to the abundance of publicly accessible data [30].

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 respective registration dates to identify those developers who had 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 collected data on other projects that each developer has participated in before to identify their past collaborators. Based on this

data, we are able to construct affiliation matrices of developers and projects that depict the existence and strengths of the relationships ties between developers.

The following measures were computed for empirical analysis (see Table 1).

Table 1. Summary of Measures

Variable	Definition
Dependent Variable	
DevelopersJoin	1 if at least one developer joined the project within the first month of project initiation, 0 otherwise.
Independent Variable	
InitiatorHasTies	1 if project initiator(s) have preexisting social ties in the network, 0 otherwise.
InitiatorTiesAmount	The amount of direct ties that the project initiators have prior to project registration calculated as the number of distinct developers who have collaborated with the project initiator(s).
InitiatorExperience	Number of projects that the project initiators have participated in before.
Control Variables	
NumInitiators	Number of project initiators[2].
ProjAmbiguity	Level of ambiguity of project definition (i.e., how ill-defined a project is) calculated as the number of project characteristics left undefined[3].

3.2 Results

The descriptive statistics and pairwise correlations of the measures for the sample are summarized in Table 2. The highest correlation between the independent variables is between *InitiatorHasTies* and *InitiatorTiesAmount* ($\rho = 0.333, p < 0.001$). The sample projects had 1.13 initiators on average. Within the first month 43% of the 1030 projects had at least one developer joining the development team. Most projects (55%) attracted one developer, 170 projects (40%) had added two to five developers, and 20 projects (5%) had more than five additional developers.

[2] The granularity of data collection is daily. In other words, we were unable to distinguish between initiators and subsequent joiners if the project registration and the developer's join event happened on the same day. We classified all members that joined on the day of registration as initiators.

[3] On SourceForge.net, project administrators may clarify the details of the project in terms of several characteristics such as development status, database environment, intended audience, license type, operating system, programming language, software category, translations and nature of user interface.

Table 2. Descriptive Statistics

Variable	Descriptive Statistics				Correlations				
	Mean	St. Dev	Min	Max	(1)	(2)	(3)	(4)	(5)
(1) DeveloerJoin	0.43	0.495	0.00	1.00					
(2) InitatorHasTies	0.26	0.441	0.00	1.00	0.06*				
(3) InitiatorTieAmount	4.01	20.149	0.00	330.00	0.09***	0.33***			
(4) InitiatorExperience	1.12	4.708	0.00	81.00	-0.04	0.33***	0.22***		
(5) NumInitiators	1.13	0.510	1.00	9.00	-0.01	0.18***	0.13***	0.30***	
(6) ProjAmbuguity	3.92	3.216	0.00	7.00	-0.22***	-0.05***	-0.05*	0.00	-0.09***

Note: Sample size N = 1030.
Significance Levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Since our dependent measure (i.e., *DevelopersJoin*) is binary, we test the impact of the existence of initiators' prior social ties on developer joining behavior (hypothesis H1) by estimating the parameters for the following logistic regression model:

$$\text{logit}\left(P\left(DevelopersJoin = 1\right)\right) = \alpha + \beta_1 InitiatorHasTies + \beta_2 NumInitiators + \beta_3 InitiatorExperience + \beta_4 ProjAmbiguity + \varepsilon$$

A positive and significant estimate of parameter β_1 would indicate that the probability of other developers becoming members of a project whose initiators have direct social ties is greater than that of a project whose initiators have no direct social ties in the network. The results of the logistic regression are presented in Table 3 (Model 1). The model shows a good fit with the data (likelihood ratio $\chi^2 = 58.428$, $p < 0.01$). The variable *InitiatorHasTies* has a significant positive effect on the likelihood of developers joining ($\beta_1 = 0.389$, $p < 0.05$). The results suggest that projects with initiators who have preexisting ties with the developer network are 47.6% more likely to have at least one additional developer join the project team compared to those with initiators who do not have any preexisting ties with the network (H1 supported).

Next we test the impact of number of prior social ties on developer joining behavior (hypothesis H2) by estimating the parameters for the following logistic regression model:

$$\text{logit}\left(P\left(DevelopersJoin = 1\right)\right) = \alpha + \beta_1 InitiatorTieAmount + \beta_2 NumInitiators + \beta_3 InitiatorExperience + \beta_4 ProjAmbiguity + \varepsilon$$

The results of the logistic regression are presented in Table 3 (Model 2). The model shows good fit with the data (likelihood ratio $\chi^2 = 24.556$, $p < 0.01$). *InitiatorTieAmount* has a significant positive effect on the likelihood of developers joining ($\beta_1 = 0.0145$, $p < 0.05$). The results suggest that an additional tie for an initiator increases the likelihood of at least one developer joining the project team by 1.5%. Given that on average an initiator has had prior relationships with approximately 4 other developers, this would amount to an average increase in the likelihood by 6%. Thus, projects with initiators with more ties with the developer network are more likely to attract additional developers than those with initiators with fewer direct ties (H2 supported).

Finally, we examined the impact of initiators' experience with open source software development projects on developer joining behavior for those projects without preexisting social ties (hypothesis H3). We estimate the parameters for the following logistic regression model:

$$\text{logit}\left(P\left(DevelopersJoin = 1\right)\right) = \alpha + \beta_1 InitiatorExperience + \beta_2 NumInitiators + \beta_3 ProjAmbiguity + \varepsilon$$

Table 3 (Model 3) summarizes the results of the logistic regression. The model shows a good fit with the data (likelihood ratio $\chi^2 = 51.092$, $p < 0.01$). The parameter estimate for *InitiatorExperience* is significant but negative ($\beta_1 = -0.604$, $p < 0.01$), indicating that projects whose initiators have more OSSD experience are less likely to attract additional developers than those whose initiators have less OSSD experience, a result which may seem counter-intuitive. An alternative explanation may be that developers in the OSS community support newcomers by joining their projects and at the same time expand their existing social relationships. Therefore, hypothesis H3 that for those projects without preexisting strong social ties the experience of initiators tend to have a positive impact on the probability of having other developers join the project team was not confirmed by the results.

Table 3. Logistic Regression Results

	Model 1 (H1)		Model 2 (H2)		Model 3 (H3)	
Variable	**Parameter Estimate**	**Odds Ratio**	**Parameter Estimate**	**Odds Ratio**	**Parameter Estimate**	**Odds Ratio**
Constant	0.2878		0.4896*		0.4986*	
InitiatorHasTies	0.3891**	1.476				
InitatorTieAmount			0.0145**	1.015		
InitiatorExperience	-0.0464	0.955	-0.0363	0.964	-0.6040***	0.547
NumInitiators	-0.0742	0.928	-0.0997	0.905	-0.1787	0.836
ProjAmbiguity	-0.1416***	0.868	-0.1400***	0.869	-0.1432***	0.867
Model Statistics						
Sample Size (*N*)	1030		271		759	
Likelihood Ratio (χ2)	58.428***		24.556***		51.092***	
Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$						

6 Conclusion and Discussions

In this study we investigated the role of social ties in OSSD team formation. Specifically, we examined whether the existence of prior social ties impacts the probability of an OSS project to attract more developers. We find that overall the existence of prior social ties does increase the probability that developers join a project. We also find that, for projects with preexisting social ties, the number of such ties has a positive influence on whether additional developers join the project. Interestingly, for projects without preexisting social ties, developers tend to join the project initiated by people with less OSSD experience. This research fills a gap in the

open source literature by conducting an empirical investigation of the role of social relations on project team formation behavior. 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.

However, the study has some limitations. For example, we only look at joining behavior within the first month after project registration. The joining behavior may differ during different stages of project development. While controlling for development stage would shed more theoretical insights, practically many newly registered projects do not define their development stages explicitly, which limits our ability to incorporate this factor into the analysis. Moreover, we assume that developers who have collaborated on a project before have developed direct social ties of uniform strength. In reality, the strength of the tie may depend on many factors such as developers' roles, duration of collaboration, and outcome of the collaboration. We hope to distinguish the strength of social ties in a follow-up study. An important extension of this paper is to study the effect of developer joining behavior on the network structural characteristics within project team as well as its performance implications.

REFERENCES

[1] E. S. Raymond and B. Young, *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary* (O'Reilly & Associates, Sebastopol, CA, 2001).

[2] T. O'Reilly, Lessons from Open-Source Software Development, *Comm. ACM.* **42**(4), 33-37 (1999).

[3] E. S. Raymond. The Cathedral and the Bazaar; http://www.4linux.com.br/arquivos/cathedral-bazaar.pdf

[4] J. Lerner and J. Tirole, The Open Source Movement: Key Research Questions, *European Econom. Rev.* **45**(4-6), 819-826 (2001).

[5] W. W. Powell, K. W. Koput, and L. Smith-Doerr, Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology, *Admin. Sci. Quart.* **41**(1), 116-145 (1996).

[6] R. Gulati, Social Structure and Alliance Formation Pattern: A Longitudinal Analysis, *Admin. Sci. Quart.* **40**, 619-652 (1995).

[7] J. Lerner and J. Tirole, Some Simple Economics of Open Source, *J. Industrial Econom.* **50**(2), 197-234 (2002).

[8] A. Hars and S. Qu, Working for Free? Motivations for Participating in Open-Source Projects, *Internat. J. of Electronic Commerce.* **6**(3), 25-39 (2002).

[9] G. Hertel, S. Niedner, and S. Herrmann, Motivation of Software Developers in Open Source Projects: An Internet-Based Survey of Contributors to the Linux Kernel, *Res. Policy.* **32**(7), 1159-1177 (2003).

[10] K. R. Lakhani and R. Wolf, in Perspectives on Free and Open Source Software, edited by J. Feller, B. Fitzgerald, S. Hissam, and K.R. Lakhani (MIT Press: Cambridge, MA. 2005).

[11] P. J. Hinds, K. M. Carley, D. Krackhardt, and D. Wholey, Choosing Work Group Members: Balancing Similarity, Competence, and Familiarity, *Organ. Behavior and Human Decision Processes.* **81**(2), 226-251 (2000).

[12] M. Granovetter, The Strength of Weak Ties, *Amer. J. Sociology.* **78**, 1360–1380 (1973).

[13] S. Wasserman and J. Galaskiewicz, *Advances in Social Network Analysis* (Sage, Thousand Oaks, CA, 1994)

[14] B. Wellman and S. D. Berkowitz, *Social Structures: A Network Approach* (Cambridge University Press, Cambridge, 1998)

[15] K. S. Cook and J. M. Whitmeyer, Two Approaches to Social Structure: Exchange Theory and Network Analysis, *Ann. Rev. Sociology.* **18**, 109-127 (1992).

[16] S. P. Borgatti and P. C. Foster, The Network Paradigm in Organizational Research: A Review and Typology, *J. Management.* **29**(6), 991-1013 (2003).

[17] W. Tsai, Knowledge Transfer in Intraorganizational Networks: Effects of Network Position and Absorptive Capacity on Business Unit Innovation and Performance, *Acad. Management J.* **44**(5), 996-1004 (2001).

[18] J. Singh, Collaborative Networks as Determinants of Knowledge Diffusion Patterns, *Management Sci.* **51**(5), 756-770 (2005).

[19] S. Huang and G. DeSanctis. Mobilizing Informational Social Capital in Cyber Space: Online Social Network Structural Properties and Knowledge Sharing. *Proceedings of the 26th International Conference on Information Systems (ICIS 2005).* Las Vegas, NV.

[20] J. Walker, S. Wasserman, and B. Wellman, in Advances in Social Network Analysis, edited by S. Wasserman and J. Galaskiewicz (Sage, Thousand Oaks, CA. 1994).

[21] R. Burt, *Structural Holes: The Social Structure of Competition* (Harvard University Press, Cambridge, MA, 1992)

[22] E. von Hippel and G. von Krogh, Open Source Software and the 'Private-Collective' Innovation Model: Issues for Organization Science, *Organ. Sci.* **14**(2), 209-223 (2003).

[23] G. Madey, V. Freeh, and R. Tynan. The Open Source Software Development Phenomenon: An Analysis Based on Social Network Theory. *Proceedings of 8th Americas Conference on Information Systems (AMCIS 2002).* Dallas, Texas.

[24] L. Lopez-Fernandez, G. Robles, and J. M. Gonzalez-Barahona. Applying Social Network Analysis to the Information in CVS Repositories. *Proceedings of the 1st International Workshop on Mining Software Repositories (MSR 2004).* Edinburgh, UK.

[25] J. Xu, Y. Gao, S. Christley, and G. madey. A Topological Analysis of the Open Source Software Development Community. *Proceedings of 38th Hawaii International Conference on System Sciences (HICSS 2005).* Hawaii, HI.

[26] N. Ducheneaut, Socialization in an Open Source Software Community: A Socio-Technical Analysis, *Computer Supported Cooperative Work.* **14**, 323-368 (2005).

[27] D. McClelland, J. Atkinson, R. Clark, and A. Lowell, *The Achievement Motive* (Appleton-Century-Crofts, New York, 1953)

[28] S. Schachter, *The Psychology of Affiliation* (Stanford University Press, Stanford, CA, 1959)

[29] D. H. Gruenfeld, E. A. Mannix, K. Y. Williams, and M. A. Neale, Group Composition and Decision Making: How Member Familiarity and Information Distribution Affect Process and Performance, *Organ. Behavior and Human Decision Processes*. **67**(1), 1-15 (1996).

[30] J. Howison and K. Crowston. The Perils and Pitfalls of Mining Sourceforge. *Proceedings of the 1st International Workshop on Mining Software Repositories (MSR 2004)*. Edinburgh, UK.