

Association for Information Systems AIS Electronic Library (AISeL)

ICIS 2009 Proceedings

International Conference on Information Systems
(ICIS)

2009

Discovering Determinants of Project Participation in an Open Source Social Network

Daning Hu

University of Arizona, daningu@cityu.edu.hk

J. Leon Zhao

City University of Hong Kong, jlzhao@cityu.edu.hk

Follow this and additional works at: <http://aisel.aisnet.org/icis2009>

Recommended Citation

Hu, Daning and Zhao, J. Leon, "Discovering Determinants of Project Participation in an Open Source Social Network" (2009). *ICIS 2009 Proceedings*. 16.

<http://aisel.aisnet.org/icis2009/16>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

DISCOVERING DETERMINANTS OF PROJECT PARTICIPATION IN AN OPEN SOURCE SOCIAL NETWORK

Completed Research Paper

Danling Hu

1.Department of Management Information Systems, Eller College of Management University of Arizona, Tucson, AZ 85721

2.Department of Information Systems, City University of Hong Kong, Kowloon, Hong Kong SAR, China
daninghu@cityu.edu.hk

J. Leon Zhao

1.Department of Information Systems, City University of Hong Kong, Kowloon, Hong Kong SAR, China

2.Department of Management Information Systems, Eller College of Management University of Arizona, Tucson, AZ 85721
jlzhao@cityu.edu.hk

Abstract

Successful open source software projects often require a steady supply of self motivated software developers. However, little work has been done from a relational/network perspective to study the factors that drive the developers to participate in OSS projects. In this paper, we investigate the participation dynamics in a social network, particularly in an online open source community called Ohloh. Through a REST-based API, we collected information about 11,530 open source software projects involving 94,330 developers. Using social network analysis and statistical analysis methods, we examine a set of social and technical factors in the Ohloh dataset, which we define as the determinants that significantly influence the developers' participation choices. We found that the determinants include (1)homophily in programming language, (2)project mutual acquaintance, and (3)project age. In addition, our research findings provide the possibility of predicting developers' participation choices based on the discovered determinants, and therefore can have important implications for OSS project management and in designing social network enabled recommendation systems.

Keywords: Participation, determinant, open source software, social network analysis, link prediction

Introduction

Recently, studies on open source have proliferated in information systems research as open source becomes a major approach to software development supported by major companies such as IBM. In addition, the use of online community for the virtual collaborations among open source developers has provided rich data sources for studying open source phenomenon. One main line of inquiry is to study why developers participate in various open source software (OSS) projects. Such studies usually focused on the factors that significantly influence the developers' project participation choices. In this study, these factors are defined as the determinants of project participation. These determinants may be technical factors (e.g., project performance) or social factors (e.g., developer's individual attributes).

Previously the determinants of OSS project participation were usually studied at individual level. Relational/network perspective is largely ignored. Since the OSS project participation process involves two related entities – the developer and the project, relational information between them needs to be analyzed to discover the determinants of project participation. In this study, we first collected project and developer information, and relational information such as project participations and peer evaluations from a large online open source community – Ohloh. We then used social network analysis (SNA) methods to model the participations as a network, in which the developers and projects are nodes and developers' participations in OSS projects are links.

Then the research question of our study becomes that how to quantitatively discover the determinants of participation links in this Ohloh participation network? To answer this question, we extracted a set of plausible determinants of participation links from the Ohloh dataset based on SNA and OSS literature. These possible determinants are then statistically examined using conditional (fixed-effects) logistic model (CLM) (McFadden et al. 1974).

We claim three contributions for this research. First, we studied the determinants of OSS project participations from a relational/network perspective. The knowledge of the discovered determinants can be used by researchers and practitioners to devise useful strategies or design collaborative information systems for OSS project management. Second, we develop a mechanism based on the conditional logistic analysis to predict the OSS developers' participation choice at the dyadic level. Such predictive mechanism can be used to provide customized recommendations of OSS projects for developers. Third, our study proposes a novel computational approach that combines both social network analysis and conditional logistic analysis to discover the determinants of links in social networks. This approach can be generalized to studies of social networks in other domains.

The remainder of this paper is organized as follows. In the next section, we provide a review of literature on open source participation and social network analysis. The third section introduces the Ohloh dataset and our data collection methods. Then we describe the research design and the experimental results on the determinants of OSS project participations. After that, we discuss the implications of the discovered determinants and propose a prediction mechanism based on those determinants. Finally, we conclude the paper and suggest directions for future work.

Related Work

Open source software development largely relies on the voluntary efforts of a community of software developers. A number of studies have explored why the developers to participate in OSS development, mainly from two aspects – intrinsic and extrinsic (Krishnamurthy 2006). The intrinsic motivations for participations are the social factors related to the needs satisfying the developer such as altruism and enjoyment, while extrinsic motivations are usually derived from external rewards such as positive evaluations, learning and career advancement.

One of the most cited work (Roberts et al. 2006) on OSS participation adopted the intrinsic and extrinsic perspectives in their theoretical model. This model examines how participations, performance, intrinsic and extrinsic motivations of OSS community members interrelate using empirical data from the Apache project. The results show that developers' performances were positively influenced by their participation levels. Moreover, this study also found that an extrinsic motivation factor – developers' desire to gain high community status – can lead to above average participation levels. At last, OSS developers' past performance was found to enhance their subsequent status motivation. (Bagozzi et al. 2006) surveyed 402 active members from 191 Linux User Groups (LUG) and found that the participation to LUG is positively related with the person's experience level in Linux. Another empirical study

on Linux project participation (Hertel et al. 2003) found that the intentions to improve one's own use of software and to increase career opportunities are important motives for participation.

However, these studies on determinants of OSS project participation suffer from two problems. Firstly, they studied the determinants mainly from an individual perspective rather than a relational one. A lot of relational information such as evaluations between developers is ignored. Such information may have great impacts on developers' participation choices. Secondly, the datasets used in these studies often just include one or two open source projects. The empirical findings may not be generalized to various kinds of OSS projects. These are mainly due to the lack of 1) large-scale empirical OSS related data, 2) and appropriate analytical methods.

To address these two problems, we adopted social network analysis and statistical methods to examine the factors that determine developers' choices to participate in OSS projects of a large online OSS community. The methods we used and the relevant studies are reviewed in the following sections.

Social Network Analysis

Social network analysis is originally developed by sociologist Jacob Moreno (1934) to investigate the relationship between social structures and personal psychological well-being. He also invented sociogram – a diagram of nodes and links used to represent relationships among social actors. In the early development of SNA, there are various other ad hoc studies in sociology, anthropology and psychology that adopted similar concepts and methods. Linton Freeman in his book (2004) about the development of SNA observed that a growing number of researchers have contributed to SNA in the 1960s. One of the most important research groups at that time, Harrison White and his students at Harvard University: Stanley Milgram (six degree of separation) (1967), Mark Granovetter, and Barry Wellman (1996) have elaborated and popularized SNA.

Nowadays, with the advance of the computing technologies and the availability of massive online data, social network analysis has been used to study various large-scale real-world networks, including networks in open source software communities (Crowston et al. 2003; Grewal et al. 2006; Jin et al. 2005; Madey 2002; Wagstrom et al. 2005). In recent years, there are mainly two lines of SNA studies. One focuses on the static topologies of social networks. The structural properties of the nodes and links are examined to describe and explain how network topologies affect the functions and behaviors of complex systems (Albert et al. 2002). Two types of social networks were extensively studied: 1) communication networks such as Internet (Albert et al. 1999), email (Newman et al. 2002) and phone (Abello et al. 1999) networks; and 2) collaboration networks such as co-authorship networks (Barabasi et al. 2002; Newman 2001b) and movie-actor networks (Watts et al. 1998). However, these studies usually focused on the static network topologies and often ignored the dynamic network processes such as link formation.

Therefore, another line of SNA research has thrived aiming to study various dynamic network processes and the determinants behind those processes. Such dynamic network analyses mainly use statistical methods to model different network processes. These models are then tested with empirical data to account for the structural changes of network topologies. The network processes studied include formations of friendship links (Leenders 1996), collaboration links (Lomi et al. 2006; Nerkar et al. 2005; Powell et al. 2005a), communication links such as emails (Kossinets et al. 2006) and phone calls (Palla et al. 2007), and co-offending links (Hu et al. 2008a).

Despite the different foci of these two types of SNA methods – topology analysis and statistical methods, they can be used together in studying the determinants of various relationships in social networks. More specifically speaking, topology analysis can help identify the possible determinants of network relationships, while statistical analysis can be used to further examine such determinants.

Topology Analysis

The emergence of a unique network topology is usually resulted from certain network processes which are influenced by determinants. Therefore, to study such determinants and their impacts on network processes, the topology of the network under study needs to be identified. Several SNA measures such as average degree, clustering coefficient, average path length, and degree distribution are developed to describe and distinguish different network topology models. Three major models have been employed to characterize complex networks: random graph model (Erdos et al. 1960), small-world model (Watts et al. 1998), and scale-free model (Barabasi et al. 1999). In random networks, each node has roughly the same number of links which equals to its average degree.

Clustering coefficient is usually used to determine the small-world nature of social networks. It is the probability that two nodes with a common neighbor also link to each other. A small-world network usually has a significantly larger clustering coefficient (Watts et al. 1998) than its random model counterpart, indicating a high tendency for nodes to form clusters and communities. In addition, a small-world network often has a relatively small average path length (i.e., average number of steps along the shortest paths for all possible pairs of network nodes) (Watts et al. 1998).

Degree distribution $P(k)$ is the probability distribution of a node has exactly k links. Power-law degree distribution is used to characterize scale-free networks (Wasserman et al. 1994). In such networks, a small fraction of the nodes have a large number of links while a big fraction of nodes have just a few. This scale-free topology may be caused by the newly joined nodes' preferential attachment to the existing nodes with high degrees (Albert et al. 2002).

However, while network topology analysis provides some clues for possible determinants of network processes, it lacks the capabilities to quantitatively examine such determinants. As a result, several statistical analysis methods are introduced to complement network topology analysis.

Statistical Analysis on Determinants of Link Formation Processes

Statistical analysis has been widely used to quantitatively study the effects of various network processes on network topology changes (Albert et al. 2002) and identify the determinants of those processes (Kossinets et al. 2006; Powell et al. 2005a). In those studies, network topology changes are subject to certain stochastic processes of network effects such as reciprocity, transitivity, and balance. Several statistical models have been developed based on these network effects (Frank et al. 1986; Hunter et al. 2006; Pattison et al. 2002; Snijders 2001; Wasserman et al. 1996). They were fitted to empirical data to examine which network effects account for the observed topology changes.

However, in this study, we focus on another type of statistical analysis that focused on identifying the determinants of link formation processes. Such statistical analysis is used in many domains such as organizational studies, sociology, and network analysis. For instance, in organizational studies, Beckman et al. (2004) studied inter-organizational network change by statistically examining factors that affect the firms' choices of partners. They analyzed data on alliance networks for the 300 largest U.S. firms from 1988 to 1993. The results showed that the stability of a firm's alliance network structure depends on the type of uncertainty it experienced. The greater the uncertainty that a firm faces alone, the more likely this firm will expand its alliance network.

In the sociology literature, Leenders (1996) used a continuous-time Markov model to study the determinants of link formation in a children's friendship network. The results showed that the homophily in gender (i.e. being the same gender) significantly affects the link (friendship) formation among children. The Markov model assumes that only the state of the network at time $t-1$ affects the current state (at time t). However, this assumption may not be valid for most real-world networks. Not only limited to friendship, McPherson et al. (2001) argue that various other social relationships, including marriage, work, advice are also influenced by the homophily principle - similarity breeds connection. In addition, Snijders (2001) developed a class of actor-oriented models to examine if the nodes adjust their linking choices in the network based on certain parameters such as their degrees. However, these models assume that the nodes are aware of their positions with respect to the whole network which is often not true in large complex networks.

Another study done by Powell et al. (2005b) examined the determinants of the partner choices for biotechnology firms in 1990s. These determinants include preferential attachment and homophily (i.e. people tend to interact with others having similar characteristics) factors using McFadden's (1980; 1974) discrete choice model, a variant of the conditional logistic model. This model is usually used to statistically analyze the human behavior of making choices. In this model, a subject is presented with choice alternatives and asked to choose the best alternative. In addition, the explanatory variables are alternative-specific or subject-specific. One limitation of this model is that it requires detail personal information of the subjects and the alternatives. Such information is usually very limited in many empirical data sources.

In addition, longitudinal network data were employed to study network determinants too. Kossinets and Watts (2006) used Cox survival analysis to identify determinants of the email link formation in a university campus over one year time period. They found that the mutual acquaintance (i.e. two individuals are acquainted with a common person) and shared class affiliations (i.e. attending the same class) significantly affect the future email link formation between two students. In addition, a similar survival analysis approach was also used by Nerkar and Paruchuri

(2005) to determine that if network centrality of inventors had a statistically significant effect on the intra-firm citation of their patents. Survival analysis lends itself well to the longitudinal analysis of network data since it involves the modeling of time to event data. In the context of determinant analysis, an event is the formation of a link. However, sometimes the time information is not available or inaccurate in the network datasets, which makes survival analysis not suitable for identifying determinants.

In general, most existing statistical models for analyzing the determinants of network link formation are limited by specific assumptions or data completeness issues. There is a lack of statistical techniques which are general enough and can be applied on empirical network datasets with missing information.

Social Network Analysis on OSS Community Networks

Many OSS empirical studies including the ones focusing on project participation collected data from online communities such as Sourceforge.net. There are mainly two types of OSS community analysis. The first type studies the composition and participation processes of the OSS community members (Bagozzi et al. 2006; Krishnamurthy 2006; Roberts et al. 2006). Koch et al. (2002) analyzed the logs of source code changes for an OSS project and identified a core set of developers who produce most of the source code output. Such core OSS community members are also found to have most intensive communications in a project (Roberts et al. 2006). Another study done by Von Krogh et al. (2003) found that new OSS community members gain benefits from specializing their initial contributions.

The second type of analysis aims to understand various relationships such as collaboration and communication among OSS community members. For instance, Ducheneaut (2005) observed that successful OSS developers progressively enroll into a collaboration network of human and material allies to support each other. Another descriptive study (Yutaka et al. 2000) found that the communications in OSS development heavily relies on electronic media (e.g., forum, mailing lists) rather than face-to-face contact. In addition, Bergquist et al. (2001) found that OSS community members gain trust from others by actively giving out high quality source code and answering questions.

Social network analysis, especially topology analysis, has been widely used to study OSS community networks. Madey (2002) studied a collaboration network of OSS developers in SourceForge.net and found it displays the scale-free network features. A more recent empirical analysis (Jin et al. 2005) of SourceForge data has discovered similar scale-free features in the collaboration network. The small fraction of the developers with a large number of collaboration links can be explained by people's tendency to collaborate with high-profile, skillful members. This preferential attachment factor may apply to project participation links as well in our study. Moreover, small-world network features – large clustering coefficient and small average path length – were also found in those SourceForge social networks.

In addition, social network models and theories are proposed to explain the network topologies of OSS networks. Crowston (2003) have found that the network topologies of bigger OSS projects are less centralized. This may be caused by the modularization process of these larger projects. Another study by Wagstrom et al., (2005) used data from blog links and mailing lists to simulate OSS network evolution, aiming to develop and validate a model which can explain developers' project participation choices. Moreover, Grewal et al. (2006) examined OSS collaboration network embeddedness and discovered it has more influence on the technical success than the commercial success of OSS projects.

Although SNA has been increasingly used to analyze the structures of various OSS relationships in the above studies (Crowston et al. 2003; Grewal et al. 2006; Hu et al. 2008b; Jin et al. 2005; Koch et al. 2002; Wagstrom et al. 2005), there is little empirical investigation of the underlying determinants of OSS project participation relationship. Moreover, previous network studies on OSS related relationships focus on the group level rather than the dyadic level. For instance, in a group level analysis, it was often assumed a group of developers evaluate a project with the same status score. However, this approach tends to ignore the heterogeneity of participation behaviors among different developers. In this study, a dyadic level prediction mechanism is developed to reflect a developer's unique opinion on a project.

This article departs from the previous literature on OSS network analysis in three ways. First, this study focuses on the determinants of participation link formation rather than network topology. Second, the proposed computational approach combined both SNA topology analysis and statistical methods while the latter is usually missed in existing

OSS community network analysis. Third, a mechanism is developed to predict the OSS developers’ participation choice at the dyadic level.

Dataset

The dataset for this study was collected from a large online OSS community – Ohloh through its API (Allen et al. 2009), which provides information about 11,530 OSS projects involving 94,330 developers. This data source is unique comparing with other major OSS communities such as SourceForge.net mainly in two ways. Firstly, it not only provides OSS project participation information but also evaluation information among Ohloh community members. Each Ohloh member can send any other member a link called “Kudo” which is a simple gesture of thanks, praise, or endorsement. Sometimes a “Kudo” link can be given to a co-developer in the same OSS project as positive evaluation for his or her contribution. Sometimes people receive “Kudo” links from others as recognition of their programming skills or appreciation for their help. Therefore, the “Kudo” evaluation links may indicate many underlying social relationships among OSS community members. Moreover, Ohloh provides information about the attributes of registered developers and projects while Souceforge.net does not. These information include developers’ attributes such as nationalities and locations, and projects statistics like total lines of codes and comment ratio. Such information is crucial for the determinant analysis of social networks.

Second, Ohloh dataset covers a more comprehensive list of major OSS projects than Sourceforge.net because of its data sources. It retrieves OSS related data from three major software revision control repositories – Subversion, CVS and Git while SourceForge.net only has data from Subversion.

In addition, Ohloh website provides several other types of information about OSS projects. For example, the project activity information keeps track of every change made about an OSS project, including what was changed, when it was changed, and who made the change. Other global statistics like programming language usage are also included. Such information coupled with the results from social network analysis may provide insights about the determinants of link formations in Ohloh networks.

Data Collection and Preprocessing

Ohloh web site provides a REST-based application programming interface (API) for users to access and query its data. We developed a set of Java programs to automatically query and retrieve OSS related data using this Ohloh API. Figure 1 shows sample data for project Firefox retrieved through this API. Since all retrieved data items are in XML format, a customized parser program was developed to parse all the data into the Ohloh database. Figure 2 shows a sample of one Ohloh database table which stores the parsed OSS project information.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <response>
  <status>success</status>
- <result>
  - <project>
    <id>9</id>
    <name>Mozilla Firefox</name>
    <created_at>2006-10-10T15:51:31Z</created_at>
    <updated_at>2009-05-05T15:31:19Z</updated_at>
    <description>The award-winning Web browser is now faster, more secure, and fully customizable to your online life. With more than 15,000 improvements, Firefox 3 is faster, safer and smarter than ever before. If you stack this project, you should also stack the Mozilla Core.</description>
    <homepage_url>http://www.firefox.com/</homepage_url>
    <download_url>http://getfirefox.com/</download_url>
    <url_name>firefox</url_name>
    <medium_logo_url>http://s3.amazonaws.com/attachments/6346/firefox_med.png</medium_logo_url>
    <small_logo_url>http://s3.amazonaws.com/attachments/6346/firefox_small.png</small_logo_url>
    <user_count>5243</user_count>
    <average_rating>4.49937</average_rating>
    <rating_count>1586</rating_count>
    <analysis_id>535485</analysis_id>
  - <analysis>
    <id>535485</id>
    <project_id>9</project_id>
    <updated_at>2009-05-02T06:33:47Z</updated_at>
    <logged_at>2009-05-02T06:20:56Z</logged_at>
    <min_month>2002-04-01T00:00:00Z</min_month>
    <max_month>2009-04-01T00:00:00Z</max_month>
    <twelve_month_contributor_count>33</twelve_month_contributor_count>
    <total_code_lines>63683</total_code_lines>
    <main_language_id>6</main_language_id>
    <main_language_name>JavaScript</main_language_name>
  
```

Figure 1. Sample Data Retrieved from Ohloh API

Project_id	Project_name	Project_URL	# of Users	Average Rating	Rating_count
9	Mozilla Firefox	http://www.firefox.com/	2307	4.535897435	390
1	Subversion	http://subversion.tigris.org/	2094	4.43884892	278
72	Apache HTTP Server	http://httpd.apache.org	1621	4.609589041	146
3141	Linux Kernel 2.6	http://www.kernel.org/	1230	4.806666666	150

Figure 2. Sample Data Parsed into the Project Table in the Ohloh database

Research Design

To answer the research question, we proposed a set of analytical methods which combine both SNA topological analysis and conditional logistic analysis to discover determinants of participation links in the Ohloh social networks. The research design is presented in Figure 3. It consists of two steps. The first step involves two components: network construction and determinants extraction. We constructed a participation network from Ohloh dataset. Meanwhile, the potential determinants of participation links may be extracted based on literature and theoretical conjectures on social networks. The second step, network analysis, contains both SNA topological analysis and determinant analysis. The details of the design are introduced in the following sub-sections.

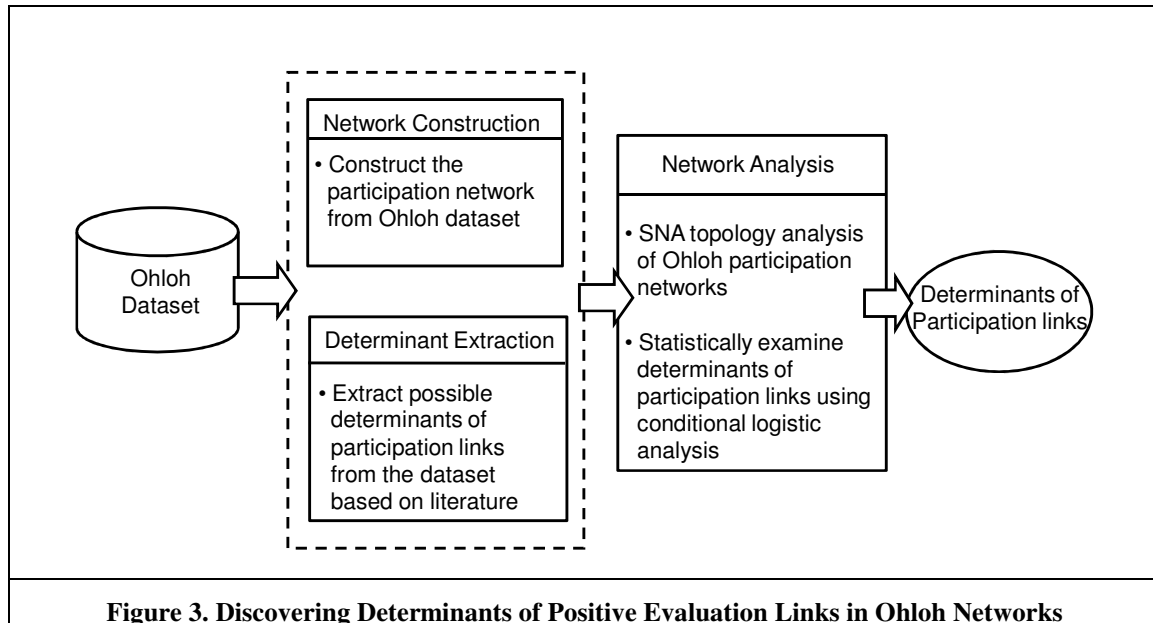
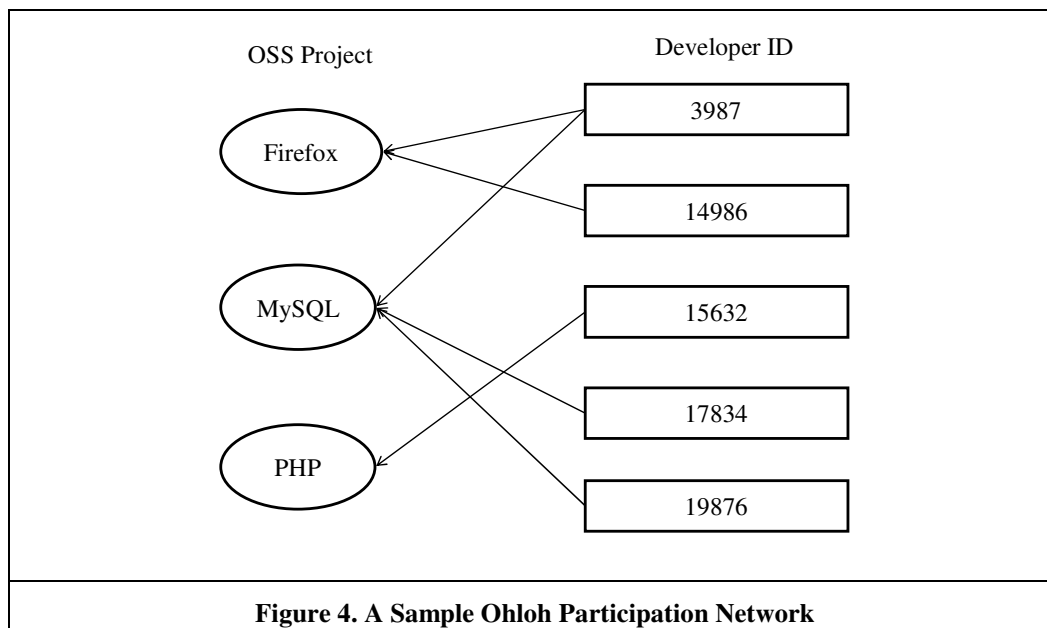


Figure 3. Discovering Determinants of Positive Evaluation Links in Ohloh Networks

Network Construction

In this section, we construct a social network by using the Ohloh participation dataset. The Ohloh participation dataset can be naturally represented as a network by treating the developers and projects as nodes, and developer's participation in projects as links. We defined such network as the participation network, which is a bipartite network (i.e. two-mode network) consisting of two sets of nodes and no links within the same set. Such bipartite networks are also termed as affiliation networks (Davis et al. 2003) in social network literature. Figure 4 shows a sample participation network from the Ohloh dataset. The round nodes represent OSS projects in Ohloh community while the square nodes are the developers. In this network, a link from developer 3987 to the project Firefox represents that developer 3987 had participated in GNOME project.



The participation dataset used in our analysis only include developers and projects which registered with complete attribute information such as developer’s nationality and project’s primary programming language. Table 1 shows the key statistics of the constructed Ohloh participation network. It contains 4690 registered developers, 5,351 OSS projects, and 11,532 participation records among them.

	Number
Developers	4690
Projects	5351
Participation Links	11532

Determinants Extraction

In this study, the possible determinants of participation were extracted from Ohloh dataset based on findings or conjectures from prior studies in both OSS domain (Roberts et al. 2006) and social network domain (Kossinets et al. 2006; McPherson et al. 2001; Powell et al. 2005b).

Project Performance Related Determinants

Previous OSS studies (Roberts et al. 2006) show that developer and project performance may significantly enhance developers’ motivation in project participation and to gain higher status. Crowston et al. (2006) applied the most commonly cited model for IS success (DeLone et al. 2003; DeLone et al. 2004) in the context of OSS development and developed a set of performance measures for OSS projects. These measures include system creation, system quality, system use and system consequences. We adopted and applied these measures in Ohloh dataset to extract the following possible participation determinants related with project performance.

System creation factors mainly measure the activity level and the overall efforts of the contributing developers on each project.

- **Total Lines of Codes:** Total lines of source code contributed by all the developers in an Ohloh project. Blank lines and comment lines are excluded.

- **Total Man Power Invested:** The cumulative total months of effort spent by all participating developers on this project. For instance, if 1 contributing developer has worked for 3 months and the other 2 have worked for 5 months, total man power invested will be 13.
- **Commits:** The total number of commits made by all contributing developers on an Ohloh project.

System quality factors measure the code and documentation quality of OSS projects, such as understandability, completeness, maintainability, usability, and efficiency (Crowston et al. 2006). In Ohloh dataset, we extracted the comment ratio for each OSS project.

- **Comment Ratio:** The fraction of total lines of code which are comments. Comments are usually very useful for the modifications and maintenance of source code. It measures the maintainability of source code.

System use factors measure usability and user satisfaction of the open source software. In Ohloh dataset, we extracted the number of users, average rating score and rating respond rate as system use factors.

- **Number of Users:** The number of users who used this OSS software project. Higher number of users indicates more popular projects.
- **Average Rating Score:** A floating point value from 1.0 to 5.0, representing the average value of all user ratings. 1.0 is the worst possible rating and 5.0 is the highest possible rating.
- **Rating Respond Rate:** The fraction of the users of an OSS project who rated that project.

SNA Related Determinants

Prior SNA studies found that the determinants of social relationships may include homophily in individual attributes and shared affiliation such as mutual acquaintance (Kossinets et al. 2006; McPherson et al. 2001; Powell et al. 2005b).

Homophily is the phenomenon that people with similar attributes are more likely to form social relations such as friendship and collaborations (McPherson et al. 2001). In Ohloh dataset, a possible determinant for participation is homophily in programming language,

- **Homophily in Programming Language:** It indicates if the contributing developer's primary programming language is the main language for the OSS project. To our best knowledge, this proposed possible determinant for the first time examines if the developers' preference on programming language influence his or her project participation choices.

Shared affiliation factors among social entities usually refer to indirect social relationships such as mutual acquaintance. They are found to be the determinants to link formation processes in various social networks (Kossinets et al. 2006).

- **Project Mutual Acquaintance with Evaluation Relationships:** It indicates if the developer has positively evaluates another developer and they both participated in the same OSS project under study. This shared affiliation factor is included because a developer may tend to participate an OSS project with many of his or her acquaintance.

Another important SNA related determinant is the preferential attachment factor (Albert et al. 2002). Preferential attachment refers to the link formation processes that nodes with certain accumulative advantage such as older age, more existing links are more likely to attract new links in the network.

- **Project age:** In Ohloh dataset, we operationalized the preferential attachment factor as the project age which is the number of the years the OSS project existed. The assumption is that projects with older age are more likely to attract new developers to participate.

Network Analysis

After we construct the Ohloh participation network and extract potential determinants, we conduct SNA topology analysis on these two networks and examine the extracted determinants using conditional logistic analysis.

Topology Analysis

There are two goals for the topology analysis of Ohloh networks in our study. First, it helps us uncover the structures of Ohloh open source software community, and better understand the nature of OSS participation relationships. Second, the determinants of link formation are found to significantly affect the network topologies (Kossinets et al. 2006; Powell et al. 2005b). For instance, the preferential attachment factor – older project – may attract more developers than younger project in the network, causing a scale-free topology in the participation network. Therefore, the results of topology analysis can be used to find plausible network determinants for further statistical analysis.

We use SNA centrality measures to describe the topology of the Ohloh participation network and identify its key members. High degrees usually indicate high levels of activity and wide social influence. The average degree of a network is also calculated to measure how dense a network is. In addition, previous research (Jin et al. 2005; Madey 2002) found that OSS collaboration networks are scale-free networks and have small-world network properties. Thus we examine the Ohloh participation network to see if it has these features. Several SNA measures are examined, including the average path length, the clustering coefficient, link density, and the degree distribution. These properties are then, checked against the small world and scale-free models.

Statistical Analysis to Examine Determinants

Our choice of the statistical model for analyzing determinants of participation in Ohloh network is based on both theoretical and empirical considerations. Theoretically, our study intends to model human (participation) choice behaviors in social networks. The research question asks how to discover determinants account for differential (as opposed to random) patterns of the link formations in OSS participation networks. Empirically, we need to model the choice behavior of a dyad - developers to projects. For these reasons, we choose to use conditional logistic model that takes each choice as a unit of analysis, which in our case is the formation of a participation link between a developer and an OSS project.

Conditional logistic model (CLM) and its variations (McFadden 1980; McFadden et al. 1974; Powell et al. 2005b) have been widely used to model human choice behaviors and examine the determinants that influence those choices. In our study, for the Ohloh participation network, the probability of an OSS developer i choose to participate an OSS project j from the alternative set J_i , is specified as follows:

$$\Pr(y_i = j) = \frac{\exp(X_j\beta)}{\sum_j \exp(X_j\beta)}$$

where y_i is the observed choice for developer i and X_j is a vector of the characteristics of the project j . The unknown coefficients β are typically estimated by maximum likelihood methods.

We estimate the CLM for the Ohloh participation network data using *clogistic* command in Stata 10/MP. The dependent variable is a binary indicator of the outcome for participation link formation between an Ohloh developer and project. The independent variables are the selected potential determinants explained in the determinant extraction section. In addition, for the statistical analysis, these independent variables are shown in Table 2.

Table 2: Variables Selected for Analysis	
Label	Description
Dependent Variable	
Participation Link	Computed as a binary indicator of whether a developer participated in an OSS project
Independent Variables	
Project Performance Factors	

Total Lines of Code	The total net lines of code, excluding comments and blanks, as of the end of this month.
Total Man Power Invested	The cumulative total months of effort expended by all contributors on this project
Commits	The cumulative total number commits to the project source control
Comment Ratio	The fraction of net lines which are comments
Number of Users	The total number of users of this open source software
Average Rating Score	A floating point value from 1.0 to 5.0, representing the average value of all user ratings. 1.0 is the worst possible rating and 5.0 is the highest.
Rating Respond Rate	The number of users who rated this open source software divided by the total number of users.
SNA Related Factors	
Homophily in Programming Language	Set to '1' if the developer's primary programming language is the same with the main language of the project, and '0' otherwise
Project Mutual Acquaintance with Evaluation Relationships	Set to N if the developer has positively evaluated (sent a Kudo link to) N (N>=1) developers in the OSS project under study
Project Age	No. of calendar months in which the OSS project existed.

Results

Topology Analysis

We start the analysis from describing the topology of the participation networks. Table 3 shows that the results of SNA topology measures for the two types of nodes in the whole Ohloh participation network. The average degree of developer nodes is the average number of projects a developer participates. On the other hand, the average degree for project nodes is average developers a project has. In addition, since prior studies (Jin et al. 2005; Madey 2002) found that OSS collaboration networks are scale-free networks, we then fit the degree distributions of the two networks to power-law distribution to test for scale-free topological features.

Table 3: Results of SNA Measures of the Ohloh Participation Network		
	Developer Nodes	Project Nodes
Number	4690	5351
Average Degree	2.60	2.28
Max Degree	55	168
Min Degree	1	1
Degree Distribution		
R^2	0.96	0.89
γ	2.27	1.81

As Table 3 shows, we found that the average degrees of these two types of nodes are similar while the max degree of project nodes is much larger than the developer nodes. This indicates there is more heterogeneity in the project nodes than the developer nodes. The last three rows of Table 3 show the results of linear regression on the degree distribution of the Ohloh participation network. It was found that this network follows the power-law degree

distribution (Newman 2001a) for both developer nodes and project nodes, $p(k) \sim k^{-\gamma}$. The coefficients of determination R^2 is extremely large at 0.96 and 0.89 (ranging from 0 to 1) respectively, indicating high fitness of the power-law degree distribution model. These findings from topology analysis imply that the Ohloh participation network is a scale-free network whose evolution may be significantly influenced by the preferential attachment factor such as project age. Therefore, we include project age as a possible determinant into the statistical analysis.

One thing to note is that most network topological models such as small-world or scale-free are developed with regard to one-mode networks. However, in our study, we are analyzing a two-mode network with two different sets of nodes – projects and developers. In this case, the power-law degree distributions for developer nodes indicate a few projects have a large number of developers while most others only have a few. Therefore, the underlying preferential attachment processes should still apply to the scale-free topology of our two-mode participation networks.

Statistical Analysis

Since the information about OSS projects is provided and updated monthly in Ohloh web site, our statistical analysis focused on developers who has participated in OSS projects in the most recent month data (May 2007). This dataset includes 229 developers who participated 772 projects in May 2007. In addition, the conditional logistic analysis requires a set of alternative projects for the developer in each participation occurrence. The number of possible participation links is the possible combinations between developers and projects, calculated as the number of developers multiplies the number of alternative projects. Since the average age of OSS project in Ohloh dataset is around 2.5 years, we limited the alternative set of projects to the ones which started within the past three years (of May 2007). Table 4 shows the descriptive statistics of the data sample used in our statistical analysis.

Table 4. Key Statistics of the Ohloh Data Sample from Conditional Logistic Analysis				
Variables	Mean	Standard Deviation	Min	Max
Total Lines of Code	181813.1	566497.5	12	7914568
Total Man Power Invested	160.105	340.8	1	6126
Commits	1323.197	2965.4	2	48560
Comment Ratio	0.211	0.105	0	0.574879
Number of Users	12.631	31.501	0	359
Average Rating Score	3.533	2.015	0	5
Rating Respond Rate	0.172	0.447	0	1
SNA Related Factors				
Homophily in Programming Language	0.13997	0.346957	0	1
Project Mutual Acquaintance with Evaluation Relationships	0.006081	0.116234	0	10
Project Age	1.398964	0.921198	0	3

To facilitate the interpretation of the results from conditional logistic analysis, we present the odds ratios and the coefficients as Table 5 shows. The odds ratios are obtained from the coefficients by using exponential function. That is $X_i = \exp(\beta_i)$ where β_i is the coefficient. In the Ohloh context, the odds ratio measures the change of the probability that a participation link is formed caused by each unit increase in an independent variable. This means that the probability of the link formation would increase by a factor of odds ratio when the corresponding

independent variable increases by one unit. Odds ratio equals to one means there is no effect of the independent variable on the dependent variable. In addition, the Pseudo R square of this regression is 0.1987.

Table 5: Results from Conditional Logistic Regression Analysis		
	Odds Ratio X_i	Coefficient β_i
Project Performance Factors		
Total Lines of Code	1.000	-2.30e-08
Total Man Power Invested	1.001*	0.002*
Commits	1.000*	-.0001*
Comment Ratio	0.492	-0.709
Number of Users	1.005*	0.006*
Average Rating Score	0.968	-0.032
Rating Respond Rate	0.971	-0.028
SNA Related Factors		
Homophily in Programming Language	8.480*	2.137*
Project Mutual Acquaintance with Evaluation Relationships	28.001*	3.332*
Project Age	0.771*	-0.260*

* $p < 0.01$

Table 5 shows the odds ratios and coefficients for each independent variables from the CLM analysis on the Ohloh participation networks. Only independent variables with p-value less than 0.05 and odds ratios significantly different from 1 are determinants. The results show that the homophily in programming language and project mutual acquaintance are determinants of participation links with odds ratios larger than 1. In addition, the project age factor is also found to be the determinant with odds ratio smaller than 1. This means that the probability for project participation decreases as the age of this project grows.

Discussion

Determinants of OSS Project Participation

As the results of the conditional logistic analysis indicated, a developer in Ohloh community is more likely to participate in an OSS project if 1) his primary programming language is the same with the project main programming language, or 2) he has positively evaluated one or more developers in that OSS project before. Moreover, a developer is less likely to participate in an OSS project as its age grows.

These findings may partly be explained by the following conjectures. First, skills in Programming language is one of the core competencies needed by OSS developers and projects (Crowston et al. 2002). OSS developers autonomously develop their skills of programming language by participating in projects. Skillful developers may tend to work in the projects that use the programming languages they are most familiar with, while junior developers tend to seize every opportunity to practice their primary programming language (Yunwen et al. 2003). In addition, some OSS project may require their participants equipped with knowledge and skills of certain programming language to maintain the quality of the development work

Another determinant – (project) mutual acquaintance with evaluation relationship - has been studied in several other SNA research (Hu et al. 2008c; Kossinets et al. 2006; McPherson et al. 2001). It was found that individuals tend to select new acquaintances who are friends of a friend. In the Ohloh participation network, this finding suggests that

OSS developers tend to form circles of close acquaintances and are attracted to OSS projects that their acquaintances are in. Another possibility is that the developers of an OSS project may tend to recruit new members from their close acquaintances. One implication of this determinant is that, recruiting developers with good reputation and high community status may also attract his close acquaintances to the same project. These close acquaintances participate in the projects usually for the opportunities to learn from or cooperate with the original developer.

Surprisingly, the preferential attachment factor – project age – was found to negatively influence the developers' project participation choices. This finding means in time the OSS projects gradually lose its capability to attract new developers. One possible explanation is that, with the rapid development of information and software technology, developers newly joined in the Ohloh community are less likely to be interested in developing older software or have the older skills required by such development.

Insignificant Factors

Another surprising finding is that all project performance factors have little effect on the developers' choices to participate in OSS projects. It may be because that the project performance information, such as total lines of codes, is often not available for potential contributing developers. In addition, for new developers with little OSS experiences, it is often difficult for them to understand such project performance information. On the other hand, social factors such as mutual acquaintance may play a more important role in attracting such inexperienced developers. One very practical implication for the OSS project managers is that, simply advertising strong project performance may not be useful in attracting and recruiting developers. Personal connections or other customized information, such as an invitation from a respected developer friend is critical for such recruiting tasks in OSS project management.

Impacts of Determinants of OSS Project Participation

In addition, the coefficients of the discovered determinants vary from 0.771 to 28.001. We found that the more personal the determinant's context, the larger the coefficient is. This means, having a mutual acquaintance with evaluation relationship is a stronger personal connection between the developer and the project, than just using the same programming language, therefore weighs more in the developer's decision in participating in the project. This finding may indicate that social and personal factors have larger impacts on OSS developers' participation choices than the project performance factors.

CLM-based Link Prediction Mechanism

In addition, the results from conditional logistic analysis can also be used to calculate the probability for a developer to participate in an OSS project. For instance, as Table 5 shows, in the Ohloh participation network, the coefficients for homophily in programming language and for having a project mutual acquaintance are 2.137 and 3.332, respectively. Assuming both developer *a* and project *b* uses Java as their main programming language. In addition, *a* has positively evaluated 2 other developers in Project *b* before. Then using the conditional logistic model, the probability for developer *a* to participate in project *b* in an alternative project set J_i , can be calculated as $\Pr(y_a = b) = \exp(1 \times 2.137 + 2 \times 3.332) / \sum_j \exp(X_j \beta)$. This calculation can be applied to any pair of developer and project in the Ohloh community, and therefore be used to predict how likely an Ohloh developer is going to participate in an OSS project.

This prediction mechanism can be used for OSS project management in two ways. Firstly, it can help OSS project managers to identify potential developers in the community and devise useful strategies to attract time. More specifically speaking, based on the homophily in programming language factor, the manager may need to put more advertising and recruiting efforts in technical communities and forums which are related to the programming language used by his project. Secondly, this mechanism can also be used to recommend inexperienced developers projects which are suitable for them.

Conclusions

Although there have been a fair amount of studies analyzing the motivations of OSS project participation, such work mainly focused on attributes of individual developer or project, but largely ignoring the relationships between these two entities and the underlying determinants that significantly affect the developers' participation choice. In this study, we used both SNA and conditional logistic analysis to study the determinants of developers' participation choices in a large online OSS community – Ohloh. It was found that the homophily in programming language, and project mutual acquaintance are significant determinants. The set of methods used in this research can also be applied to study determinants and topologies of social networks in other domains.

We also explored the possible social causes and implications for these identified determinants. Our analysis may help researchers and practitioners in OSS community to better understand the developers' participation choice behaviors and devise useful strategies for OSS project management. For instance, as stated in the discussion, the project manager may improve the recruitment by advertising in online communities and forums that are related to the programming language his project uses.

In addition, based on the identified determinants and conditional logistic model, we proposed a prediction mechanism that can quantitatively estimate Ohloh developer's project participation choice at the dyadic level. This prediction mechanism may be used in the design of collaborative information systems to support OSS development.

Our future work consists of three research directions including (1) further investigating the determinants and evolution of OSS participation networks with additional data and other advanced SNA methods, (2) exploring other factors that may positively influence developers' project participation choices, and (3) using such knowledge and the link prediction mechanisms derived from this study to design various information systems to support knowledge management and decision making.

References

- Abello, J., Pardalos, P., and Resende, M.G.C. "On maximum clique problems in very large graphs," in: *External memory algorithms*, American Mathematical Society, 1999, pp. 119-130.
- Albert, R., and Barabasi, A.L. "Statistical mechanics of complex networks," *Reviews of Modern Physics* (74:1), Jan 2002, pp 47-97.
- Albert, R., Jeong, H., and Barabasi, A.L. "Internet - Diameter of the World-Wide Web," *Nature* (401:6749) 1999, pp 130-131.
- Allen, J., Collison, S., and Luckey, R. "Ohloh Web Site API," <http://www.ohloh.net>, 2009.
- Bagozzi, R.P., and Dholakia, U.M. "Open source software user communities: A study of participation in Linux user groups," *Management Science* (52:7), Jul 2006, pp 1099-1115.
- Barabasi, A.-L., and Albert, R. "Emergence of Scaling in Random Networks," *Science* (286:5439) 1999, pp 509-512.
- Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., and Vicsek, T. "Evolution of the social network of scientific collaborations," *Physica A: Statistical Mechanics and its Applications* (311:3-4) 2002.
- Beckman, C.M., Haunschild, P.R., and Phillips, D.J. "Friends or Strangers? Firm-Specific Uncertainty, Market Uncertainty, and Network Partner Selection," *ORGANIZATION SCIENCE* (15:3), May 1, 2004 2004, pp 259-275.
- Bergquist, M., and Ljungberg, J. "The power of gifts: organizing social relationships in open source communities," *Information Systems Journal* (11:4) 2001, pp 305-320.
- Crowston, K., and Howison, J. "The social structure of open source software development teams," *OASIS 2003 Workshop (IFIP 8.2 WG)* 2003.
- Crowston, K., Howison, J., and Annabi, H. "Information systems success in free and open source software development: theory and measures," *Software Process: Improvement and Practice* (11:2) 2006.
- Crowston, K., and Scozzi, B. "Open source software projects as virtual organisations: competency rallying for software development," *Software, IEE Proceedings - (149:1)* 2002, pp 3-17.
- Davis, G.F., Yoo, M., and Baker, W.E. "The Small World of the American Corporate Elite, 1982-2001," *Strategic Organization* (1:3), August 1, 2003 2003, pp 301-326.
- DeLone, W.H., and McLean, E.R. "The DeLone and McLean model of information systems success: a ten-year update," *Journal of Management Information Systems* (19:4), Spr 2003, pp 9-30.
- DeLone, W.H., and McLean, E.R. "Measuring e-commerce success: Applying the DeLone & McLean information systems success model," *International Journal of Electronic Commerce* (9:1), Fal 2004, pp 31-47.
- Ducheneaut, N. "Socialization in an Open Source Software Community: A Socio-Technical Analysis," *Computer Supported Cooperative Work (CSCW)* (14:4) 2005, pp 323-368.
- Erdos, P., and Renyi, A. "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.* (5) 1960, pp 17-61.
- Frank, O., and Strauss, D. "Markov Graphs," *Journal of the American Statistical Association* (81:395) 1986.
- Freeman, L.C. *The Development of Social Network Analysis* Empirical Press, 2004.
- Grewal, R., Lilien, G.L., and Mallapragada, G. "Location, location, location: How network embeddedness affects project success in open source systems," *Management Science* (52:7), Jul 2006, pp 1043-1056.
- Hertel, G., Niedner, S., and Herrmann, S. "Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel," *Research Policy* (32:7) 2003, pp 1159-1177.
- Hu, D., Kaza, S., and Chen, H. "Identifying Significant Facilitators of Dark Network Evolution," (*Forthcoming*) *Journal of the American Society for Information Science and Technology* 2008a.
- Hu, D., and Zhao, J.L. "A Comparison of Evaluation Networks and Collaboration Networks in Open Source Software Communities," Americas Conference on Information Systems, Toronto, ON, Canada, 2008b.
- Hu, D., and Zhao, J.L. "Expert Recommendation Via Semantic Social Networks," International Conference on Information Systems, Paris, France, 2008c.
- Hunter, D.R., and Handcock, M.S. "Inference in Curved Exponential Family Models for Networks," *Journal of Computational & Graphical Statistics* (15) 2006, pp 565-583.
- Jin, X., Yongqin, G., Scott, C., and Gregory, M. "A Topological Analysis of the Open Source Software Development Community," in: *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 7 - Volume 07*, IEEE Computer Society, 2005.
- Koch, S., and Schneider, G. "Effort, co-operation and co-ordination in an F/OSS software project: GNOME," *Information Systems Journal* (12:1) 2002, pp 27-42.
- Kossinets, G., and Watts, D.J. "Empirical Analysis of an Evolving Social Network," *Science* (311:5757), JAN 6 2006, pp 88-90.

- Krishnamurthy, S. "On the intrinsic and extrinsic motivation of free/libre/open source (FLOSS) developers," *Knowledge, Technology, and Policy* (18:4) 2006, pp 17-39.
- Leenders, R.T.A.J. "Evolution of Friendship and Best Friendship Choices," *Journal of Mathematical Sociology* (21:1-2) 1996, pp 133-148.
- Lomi, A., and Pattison, P. "Manufacturing Relations: An Empirical Study of the Organization of Production Across Multiple Networks," *Organization Science* (17:3) 2006, pp 313-332.
- Madey, G., Freeh, V., and Tynan, R. "The Open Source Software Development Phenomenon: An Analysis Based on Social Network Theory," AMCIS, Dallas, TX, 2002, pp. 1806-1813.
- McFadden, D. "Econometric Models for Probabilistic Choice Among Products," *The Journal of Business* (53:3) 1980, pp S13-S29.
- McFadden, D., and Zarembka, P. "Conditional Logit Analysis of Qualitative Choice Behavior," in: *Frontiers in econometrics*, Academic Press, 1974, pp. 105-142.
- McPherson, M., Smith-Lovin, L., and Cook, J.M. "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* (27) 2001, pp 415-444.
- Milgram, S. "Small-World Problem," *Psychology Today* (1:1) 1967, pp 61-67.
- Moreno, J.L. *Who shall survive?* Nervous and Mental Disease Publ., 1934.
- Nerkar, A., and Paruchuri, S. "Evolution of R&D Capabilities: The Role of Knowledge Networks Within a Firm," *Management Science* (51:5) 2005, pp 771-785.
- Newman, M.E.J. "Scientific collaboration networks. I. Network Construction and fundamental results," *Physical Review E* (64) 2001a, p 06131.
- Newman, M.E.J. "The Structure of Scientific Collaboration Networks," *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* (98:2), JAN 16 2001b, pp 404-409.
- Newman, M.E.J., Forrest, S., and Balthrop, J. "Email networks and the spread of computer viruses," *Physical Review E* (66:3) 2002, p 035101.
- Palla, G., Barabasi, A.-L., and Vicsek, T. "Quantifying social group evolution," *Nature* (446:7136) 2007.
- Pattison, P., and Robins, G. "Neighborhood-Based Models For Social Networks," *Sociological Methodology* (32:1) 2002, pp 301-337.
- Powell, W.W., White, D.R., Koput, K.W., and Owen-Smith, J. "Network dynamics and field evolution," *American Journal of Sociology* (110:4), Jan 2005a, pp 1132-1205.
- Powell, W.W., White, D.R., Koput, K.W., and Owen-Smith, J. "Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences," *American Journal of Sociology* (110:4), Jan 2005b, pp 1132-1205.
- Roberts, J.A., Hann, I.H., and Slaughter, S.A. "Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects," *Management Science* (52:7), Jul 2006, pp 984-999.
- Roberts, J., Hann, I.-H., and Slaughter, S. "Communication Networks in an Open Source Software Project," in: *Open Source Systems*, 2006, pp. 297-306.
- Snijders, T.A.B. "The Statistical Evaluation of Social Network Dynamics," in: *Sociological Methodology*, M.E. Sobel and M.P. Becker (eds.), Basil Blackwell, London, 2001.
- von Krogh, G., Spaeth, S., and Lakhani, K.R. "Community, joining, and specialization in open source software innovation: a case study," *Research Policy* (32:7), Jul 2003, pp 1217-1241.
- Wagstrom, P., Herbsleb, J., and Carley, K. "A social network approach to free/open source software simulation," First International Conference on Open Source Systems, 2005, pp. 16-23.
- Wasserman, S., and Faust, K. *Social Network Analysis: Methods and Applications* Cambridge University Press, Cambridge, 1994.
- Wasserman, S., and Pattison, P. "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p," *Psychometrika* (61:3) 1996, pp 401-425.
- Watts, D.J., and Strogatz, S.H. "Collective Dynamics of 'Small-World' Networks," *Nature* (393) 1998, pp 440-442.
- Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., and Haythornthwaite, C. "Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community," *Annual Review of Sociology* (22:1) 1996, pp 213-238.
- Yunwen, Y., and Kishida, K. "Toward an understanding of the motivation of open source software developers," *Software Engineering*, 2003. Proceedings. 25th International Conference on, 2003, pp. 419-429.
- Yutaka, Y., Makoto, Y., Takeshi, S., and Toru, I. "Collaboration with Lean Media: how open-source software succeeds," in: *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, ACM, Philadelphia, Pennsylvania, United States, 2000.