How Microblogging Networks Affect Project Success of Open Source Software Development

Xuan Yang
City University of Hong Kong
sherry.teen@gmail.com

Daning Hu
University of Zurich
hdaning@ifi.uzh.ch

Davison M. Robert
City University of Hong Kong
isrobert@cityu.edu.hk

Abstract
Microblogging as an emerging social media technology is becoming increasingly popular in more and more OSS communities and forms various follower networks. However, the impacts of microblogging follower network on OSS project success are rarely studied. In this study, we adopt a social network perspective to identify and hypothesize that three microblogging network mechanisms will positively affect OSS project success through knowledge sharing, and attracting more skillful and eminent developers. Using longitudinal data from a large online OSS community called Ohloh, we empirically examine the impacts of various factors for these two type of mechanisms on the commercial and technological successes of OSS projects. We found that preferential attachment and structure hole factors are supported, while accumulative advantage factors are partially supported. Our findings may provide insights for OSS stakeholders to effectively manage microblogging techniques for achieving project success.

1. Introduction
Recent years has witnessed the great successes of many open source software (OSS) projects such as Linux, Mozilla Firefox and MySQL. Such successes has demonstrated the market legitimacy of the model for open source software development [1]. In this community-based model, a large number of developers who are geographically dispersed rely heavily on various Internet-based communication technologies for collaboration as they form project teams to contribute to OSS development. Relatively less face-to-face interactions in online OSS development environments suggest that electronic communications are crucial for efficient virtual collaboration work among OSS developers and thereby OSS project success. Prior studies about OSS project success mainly studied factors such as project participation [2], developers’ motivations [3] and interpersonal trust [4]. However, the impacts of newly flourished Web 2.0 communication technology - Microblogging have been rarely studied.

Microblogging services are becoming increasingly popular in various online collaborative communities, including OSS development communities. Follower relationship is formed when one developer follow another one. Microbloggers and followers form a unique and information communication network and mobilize development-related updates constantly. From the followers’ perspective, information and knowledge from the individuals they have intentionally chosen to follow is often perceived as relevant, useful and trustworthy [5]. From the microbloggers’ perspective, the follower network provides them with an efficient channel to publish and share information microblogger developers consider as helpful to their followers. Therefore, the follower network provides a stable communication infrastructure in OSS communications that facilitates accurate and efficient sharing and mobilizing relevant knowledge and information flow, and thereby may positively affect collaborative dynamics of virtual teams like OSS project groups. However, to the best of our knowledge, the exact microblogging-related mechanisms that enables OSS project success has not be studied in both OSS and communication literature before. Our study aims to fill this gap by conducting an empirical investigation of the microblogging follower network in a large online OSS community called Ohloh. To be more specific, we draw upon open source, social network and communication research and hypothesize two mechanisms of the microblogging follower network – preferential attachment and structural holes and examine how they facilitate OSS project success.

By offering an understanding of the influences of microblogging social networks on OSS success, we
mainly have three major contributions for this research. First, this research contributes to rising stream of studies on microblogging phenomenon and follower social networks. Second, we gain richer theoretical understanding on the impacts of microblog network on OSS project success by incorporating social network perspectives. Third, our empirical findings provide suggestions for OSS project managers and other stakeholders to better manage informal communications techniques like microblogging among developers to improve OSS project success.

The remainder of this paper is organized as follows. In the second section, we review relevant OSS studies and develop the research hypotheses. The third section describes the dataset we used and discusses research testbed. The fourth section introduces our research design. The fifth section summarizes the findings of empirical analysis. Finally, we conclude with implications.

2. Research Background and Hypotheses

2.1. Microblogging in OSS Development

OSS development, which serves as a community-based model, has been witnessed its innovation regarding its liveliness and success [6]. In one way, this new model innovates in OSS developers’ inheriting modular architectures. Also, OSS developers acquire and allocate diverse problem-solving skills from fellow developers through their participation in software development [7] via Internet-based social network. Allocating knowledge resources through digital social network techniques help developers to “design better software, anticipate potential problems, solve thorny problems, and better identify user needs during development” [6], which are essentially crucial for project OSS success. Studies about OSS project success mainly focus on its measurements [8-14] and determinants [2, 15-17]. However, while existing OSS research has examined the impacts of a wide range of social and technical factors on project success, the impacts of digital communication are still fragmented.

Microblog as a rising internet-based technology has received much greater academic attentions recently. Extant studies mainly include two streams. One is to investigate microbloggers’ motives in various areas [5, 18-19]. The other types of research discuss usefulness and functionalities of microblogging technologies [20-22]. Those more recent work has been conducted to uncover the functionalities of microblog for OSS and software engineering developers. Those studies in IS, Software Engineering and CSCW (Computer Supported Cooperative Work) area have discovered various functionalities of these social media. These studies examine critical functionalities for an emergent social media – microblog for OSS development. Along this stream of literatures, microblog has been indicated to enhance transparency and collaboration for OSS developers. Furthermore, it is suggested that microblog acts a crucial role in signaling project liveliness and success. Having conducted a series of in-depth interviews, Dabbish, L. et al. have uncovered several features from social interaction via adoption of microblog [20]. For those who volunteer to contribute to OSS projects, it is important to know what others need from them, how things could be done and what is the purpose of their code contributing activities towards to overall goal. Microblog as an emergent digital media is thus chosen by more and more OSS stakeholders to facilitate their collaborative work by transferring rich knowledge resources across intertwined social network. Along this line of literatures, the usefulness of microblog suggests that those emergent technologies could help developers better manage social relationship and further produces harmonious collaboration and work output. To sum up, the existing literatures provide several cues for Microblog’s functionalities in OSS development. However, there is little research about the impacts of microblogging and its follower network on team performance, especially in OSS projects.

2.2. A Social Network Perspective on How Microblogging Techniques Affect OSS Project Success

In recent years, social network becomes an effective means to bond dispersed actors together. As a network form carrying in communication and collaboration relationship is nurtured by more and more OSS developers [23], social network analysis (SNA) has become one of the broadly acknowledged approaches for modeling and analyzing various relationships [2, 14, 24-27]. By incorporating social network lens, we propose two important mechanisms – preferential attachment and structural hole in the microblogging follower network will influence OSS project success.

2.2.1. Preferential Attachment

Empirical work detected the existence of preferential attachment in many social networks, including networks in OSS setting [25, 28-29]. Preferential attachment [30-31] refers to the process
that nodes already having many links are more likely to attract new links. In this study, we argue that preferential attachment mechanisms of the follower network can promote OSS project success by attracting more skillful developers to participate in a project [32] and achieve success in a long run.

**H1. The success of an OSS project is positively related to the average number of followers its microblogger developers have in the follower network.**

### 2.2.2. Accumulated Advantage

The accumulated advantage phenomenon is closely linked with preferential attachment. Based on the Merton’s (i.e. the innovation of this phenomenon) study, it is indicated that in social networks, this effect allows nodes possessing accumulated advantages in certain attributes to be more likely to attract new links. [33]. In our study, we propose that eminent microblogger developers with accumulative advantages (e.g., OSS community experiences) may attract more followers and yield a stronger positive influence on their followers’ project participation decisions than can common (not eminent) developers. Project participations among those followers are promoted and thereby OSS development performance is enhanced over time. Therefore, we hypothesize that:

**H2. The success of an OSS project is positively related to the average level(s) of accumulative advantage(s) its microblogging developers have within the follower network.**

### 2.2.3. Structural Holes

The structural lens of network studies often focuses on the configuration of links [34-35]. Structural hole, which is one of the most prominent mechanisms for network structure, has received tremendous theoretical and empirical attention regarding its emphasis on brokerage or network diversity [34, 36]. For knowledge workers, structural holes theory suggests that a team member’s ability to access a variety of information and resources in a social network strongly affects a team’s performance [34, 37-38]. Such information and resources from a range diverse of sources provide network benefits that are often additive rather than overlapping to the team [34] especially when nodes are connected through weak ties. For knowledge-intensive work settings, the value of structural holes unfolds in two aspects. On one hand, structural hole has been found to positively correlate to various measurements of work performance [34, 38]. This network advantageous stresses actors’ ability to access to information from non-redundant sub-circles and entities. On the other hand, a number of organizational studies have discovered the positive effects of network structural holes on organization performances in terms of innovation [39-40] and learning [41-42]. Prior studies have discovered one of the main contributions of microblog for OSS communities is to help developers to allocate knowledge resources from their fellow developers. This process is implemented well as developers are connected through networked relationship.

Based on recent work studying functionalities and usefulness of microblog in OSS community, we suggest the structural holes in follower networks affect OSS project groups in the following way. An OSS project member may gain innovative ideas or useful resources from individuals outside his close circle (often project members). This is because such individuals often possess information and resources rather different from what the group members already had and thereby may facilitate innovations. Alternative stated, an OSS project member may allocate knowledge and information flow from structurally diverse structure by connecting with disconnected circles and parties. With access to diverse knowledge resources, developers shape stronger skills for solving development problems [6], designing better software (i.e. identify users’ needs) [20]. As suggested, connection with various social circles allows developers to reveal more visibility of their own developers, which suggest larger audiences for OSS projects.

Therefore, we follow the findings of related organizational studies on structural holes and hypothesize that:

**H3. The success of an OSS project is positively related to the level of its members’ access to structural holes in the microblogging follower network.**

### 3. Research Testbed

The dataset from our study is collected from a large online OSS community – Ohloh – through its API. Ohloh, which is an online community, offers free public directory for OSS development. It also gathers a variety of information about OSS projects and developers from public software version control systems. Concurrent versions system (CVS) and Subversion are two commonly mentioned version control systems. Furthermore, it collects developers’ and projects’ activity information and keeps traces for every change in the source code of each listed OSS project. With such information, Ohloh calculates software metrics for each OSS project such
as the total number of source code changes (i.e., commits). Developer can register, edit and discuss various OSS projects in this online social-networking community. In 2009, the influences of Ohloh has been further recognized that it was reported as the largest directory for open source software (OSS) projects [43].

The information we extracted from this dataset to construct follower networks [32] is described in detail as followed:

- An Ohloh developer can choose to subscribe (follow) any microblogger in whom he or she is interested.
- All the microblog messages published by microbloggers will be included in the follower’s news feed.

A set of Java programs was developed to automatically query and used for retrieving Ohloh related data through the API. As the data items we retrieved are all in XML format, a parser program was developed to parse all the data into a database.

4. Research Design

In this study, statistical (regression) techniques for data analysis is adopted to examine the impacts of the two types hypothesized network factors – preferential attachment factors, accumulative advantages and structural holes factors – on OSS project success. We then extracted several control factors from Ohloh dataset based on previous research on OSS project success. Network measures are then derived. Dependent and independent variables constructed are introduced using data from Ohloh OSS community.

4.1. Dependent Variables

We constructed two dependent variables using Ohloh data based on existing literatures [12, 14, 44-45] to measure OSS project success. Previous literatures in OSS studies has indicated that OSS success mainly contain two parts – commercial success and technical success. We operationalized the measure of OSS success with technical success (measured by commits) and commercial success (measured by downloads, i.e. number of users) as two important criteria for OSS project success. Inclusion of these two criteria allows us to capture the effects of follower network mechanisms on both technical-oriented success and market-oriented success.

To assess the technical success of an OSS project, commits are used to reflect meaningful changes to the source code of an OSS project development. Following previous researchers [2], we then treat the amount of commits as the indicator of technical success.

We use downloads of an OSS project to measure its commercial success in that downloads indicates the popularity of the focal project, which reflect market-oriented success for OSS project [14].

These two dependent variables are operationalized as follows:

- Commits: Total number of commits that a project has until the time of data collection. The number of commits in a way reflects the amount of OSS contribution for a project.
- Downloads: The number of Ohloh community members who download this OSS software.

4.2. Independent Variables

As suggested earlier, we mainly constructed two types of network-related factors as independent variables for our analysis – the preferential attachment factors and the structural hole factors - along with a set of control variables available in our dataset as the independent variables (note: some of derived variables is developed based on our prior work [32]).

4.2.1. Preferential Attachment Factors

Average project microblogger in-degree, denoted as \( \text{In-degree} \): It is calculated as the average number of followers the microbloggers in an OSS project has.

4.2.2. Accumulated Advantage Factors

Average OSS community experience (Months): It is calculated as the average number of months per microblogger contributing for OSS development.

Microblog message counts (Tags): It is calculated as the number of microblog messages which contain a specific project’s name. Thus, more message counts may indicate a project’s better publicity.

Average updates counts (Updates): It refers to the average number of messages published by the microbloggers of an OSS project. This measurement may indicate the level of microblogging activity in an OSS project.

Follower network age (Age): It is calculated as the average length of microblogging activities per microblogger in an OSS project.

4.2.3. Structural Hole Factor(s)

Structural hole related independent variable is quantified by using “the access to brokers through
“structural holes” measure. The access to structural holes is then calculated as one minus a node’s network constraint. Network constraint measure developed by Burt [34] is used. The access to structural holes is then calculated as one minus a node’s network constraint.

4.3. Control Variables

Average Time Efforts: Average time efforts for an OSS project is measured by the average cumulative months of effort expended by developers within this project. It is an important indicator representing OSS knowledge workers’ (i.e. developers) capital.

Number of Developers: Project size is a commonly used control variable on prior literatures studying OSS project success. It is measured by the number of developers (i.e., project size) affiliated with an OSS project.

Comments: It is calculated as the total lines of comments for an OSS project.

Total Code Lines: It is calculated as the total lines of code provided by developers for an OSS project.

Comment Ratio: It is calculated as the fraction of lines which are comments for an OSS project.

Number of Rater: It is calculated as the number of raters who give rating score to an OSS project.

5. Results

5.1. Statistical Analysis

Linear regression analysis is used to examine the impacts of above defined independent variables on OSS project success. Initial analysis revealed that dependent and some independent variables were not normally distributed. In this case, linear regression analysis may yield based parameter estimates which cannot be easily interpreted. Suggested by Gelman and Hill [46], logarithmic transformation is performed on the dependent, and no-normally distributed independent variables.

<table>
<thead>
<tr>
<th>Table 1. Regression Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
</tr>
<tr>
<td>Updates</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Structural Hole</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
</tr>
<tr>
<td>Average Time Efforts</td>
</tr>
<tr>
<td>Number of Developers</td>
</tr>
<tr>
<td>Comments</td>
</tr>
<tr>
<td>Total Code Lines</td>
</tr>
<tr>
<td>Comment Ratio</td>
</tr>
<tr>
<td>Number of Rater</td>
</tr>
</tbody>
</table>

Table 1 shows the results of our statistical analysis on the factors that may significantly affect Ohloh project performance. The first hypothesis is supported for commercial OSS project success but not technological success. The commercial success of an OSS project is positively related to the average number of followers its microblogger developers have. We conjecture this may be because the preferential attachment factor (i.e., average project microblogger in-degree) is important in promoting the publicity of the underlying OSS project, but not good at attracting more contributions.

The second hypothesis is partially supported for both the technical and commercial project success. The number of tags in messages referring an OSS project will positively affect the success of this project. Moreover, the cumulative tenure (i.e. Age) and intensity for microblogging activities of project members have also been found to positively related to both commercial and technical success for OSS projects.

On the other hand, the average time invested by microblogger in an OSS project is found to be negatively related with both project successes. We conjecture several possible causes for this result. One of them is that the less experience may indicate this developer is rather young and more familiar with microblogging since it is a rather novel technology just emerged in recent years and more popular in younger people. Such advantages may improve the positive marketing effects of microblogging on OSS project success.

Another possible explanation is the maturity of OSS projects. An OSS project may achieve its summits in the production and mature stage in their development phase (development phases: planning, pre-alpha, alpha, beta, production, mature and inactive). Some projects having developers with rich
community experience in the OSS community, however, may have already arrived to the inactive stage (i.e. little code contribution and maintenance, limited size of users and potential users).

Consistent with our Hypothesis 3, we find an OSS project’s developers’ ability to access to structural holes has a positive effect on OSS project success on both technical and commercial facets. The results indicated that an OSS team member’s ability to access to structural holes in the microblogging follower network could facilitate information retrieving and resource sharing, and thereby improve project performance.

6. Discussion

In summary, the results in Table 1 has indicated that the structural hole, microblog tags counts, microblog updates counts, microblogging activity tenure (measure by follower network age) has been found to positively affect OSS project success in both commits and downloads success measures. The average number of follower links owned by developers in an OSS project will primarily affect market-oriented success (measured by downloads) for an OSS project rather than code contributions (i.e. technical-oriented success, measured by commits) for the project. Further discussion is presented in the next session.

These findings may be partly explained by the following conjectures. First, developers brokering diverse network structure may identify problems and acquire rich technical skills for solving development problems [6, 20] and identify users’ need with relative ease [20-21] through rich knowledge resources mobilized in follower network. This allows developers to contribute actively and derive user-welcomed OSS public goods (i.e. software).

Second, developers’ microblogging tenure indicates more cumulative experiences and advantages on microblogging activities. Those developers may have more knowledge on how to improve project success by using microblog techniques. Third, tagging the name of an OSS project will attract more attention by linking relevant microblogging updates to it. This highlight enables wide range of audiences to know what a project need, and how others could contribute to it. By specifying the relevance of microblogging updates to a certain project, a project may be exposed to larger size of audiences, thus attract greater number of users and more developers to participate. Fourth, the number of microblogging updates may signal the liveliness of an OSS project, which essentially serves to project success. This suggests regularly deliver updates on microblog may also drive OSS project success.

Furthermore, the effects of preferential attachments varied on commercial and technical facets of OSS project success. We have discovered that OSS projects with developers’ having more followers tend to attract more users who drive commercial success (i.e. market-oriented success). However, we find no statistically significant relationship between preferential attachment measure (i.e. indegree in follower network) and technical project success. One explanation is that, microblogging as a digital service facilitating communication and collaboration, while promote the popularity of OSS projects, may not directly provide an interactive environment for developers to produce code contribution.

Surprisingly, one of the accumulative advantage factor – Community Experiences for developers in an OSS project has been found to negatively influence the project success on both commercial and technical facets. This findings suggests in time the OSS projects gradually lose its power to attract attention including a stable increasing number of users and sustainably contributing developers [47]. One possible explanation is the maturity of OSS projects. A certain number of OSS projects in our dataset may have already arrived in “inactive” stage on development phase. Alternatively, as a project’s developers’ number of months (i.e. community experience) increase over time, the project may not similarly attract more code contributions and users.

We then plan to control for development phase in the future and investigate how the impacts of follower network mechanisms vary in different development phases.

We intend to provide several suggestions based on the above findings on the effects of microblogging follower networks on project success. First, developers with great experiences on OSS development may not drive OSS technical project success. An OSS project may incorporate “fresh blood” when stakeholders devise team composition strategies. It may be important for a project to have some new members who just started their OSS development activities but have good microblogging skills to improve its exposure in the community.

Our study also provides practical implications for OSS stakeholders to use microblogging technique for improving project success. OSS stakeholders could encourage its developers to access to a wide range of external knowledge resources through diverse structural network to acquire rich development-related knowledge resources. To attract more audience including developers and users, OSS
stakeholders may need to advertise their OSS projects by using microblogging. Developers may also be encouraged to use microblogging technique earlier to cumulate environmental richness (e.g., who knows what, who need what, who can do what) on OSS development.

7. Concluding Remarks

In this study, we investigated the impacts of the microblogging follower network on OSS project success. We take social network lens to develop theoretically driven hypotheses on how two network mechanisms affect project success on both market and technical facets. We statistically tested these hypotheses and capture the specific features of microblogging follower network on project success.

We claim several contributions. First, we contribute to the literatures on the drivers of OSS project success [2, 6, 48]. Secondly, we add value to OSS research that are studied via social networks lens [2, 6, 23, 25, 48]. Thirdly, the studies have showed the role of an emerging web 2.0 technology, which is useful for virtual team collaboration and communication if effectively managed. Specific roles of preferential attachment, accumulated advantages and structural holes are identified. Finally, the implications are discussed to OSS project managers to better manage microblog communications techniques to achieve project success.

Our future research includes studying more underlying rationales on the varied effects of network mechanisms indicators on OSS performance. Furthermore, more control variables need to be included to uncover a broader view for devising strategies (e.g. project age may be a contingency factor that the follower network mechanisms may affect project success differently in different OSS development phase).

8. References


