Mahdi Moqri, Xiaowei Mei, Liangfei Qiu & Subhajyoti Bandyopadhyay (2018) Effect of “Following” on Contributions to Open Source Communities, Journal of Management Information Systems, 35:4, 1188-1217.
This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Management Information Systems on 17 Dec 2018 (published online), available at: http://www.tandfonline.com/10.1080/07421222.2018.1523605

The Effect of “Following” on Contributions to Open Source Communities

ABSTRACT

In this study, we estimate the effect of “online following,” a basic form of online social interaction, on members’ contributions in open source software (OSS) communities, using a unique longitudinal data set containing information on over 4 million OSS developers and their social interactions over 7 years. We find that obtaining new followers in the previous month has a significant positive effect on developers’ level of contribution in the current month. The effect carries over to the next month although the marginal effect decreases. We further find that the effect of new followers on their contribution level is much stronger for freelancers than those with company affiliation. In contrast to the previous literature that posited the existence of non-monetary incentives for developers who contribute to the OSS communities, our result suggests the existence of incentives that are tied to future monetary rewards for developers on these platforms. Our findings have important implications for the OSS platforms as well as the OSS community. OSS platform designers may consult our results to learn about the social features that affect members’ contribution. We also encourage OSS community to use the “following” feature more prominently on OSS platforms to incentivize higher contribution levels to the projects.

Keywords: Open Source Software, Open Source Communities, Online Communities, Social Interaction, Online Following

The Effect of “Following” on Contributions to Open Source Communities

Introduction

Over the last decade, multiple studies have tried to answer one of the most fundamental questions in open source software (OSS) development formulated by Lerner and Tirole [33]: “Why would thousands of top-notch software developers contribute for free to the creation of a public good?” The extant research has identified three categories of motivations for developers to contribute to OSS projects: a) intrinsic motivations, such as a sense of enjoyment or accomplishment in the performance of the task, which are linked to the satisfaction of human needs for autonomy and competence; b) extrinsic motivations, such as reputation, status, and monetary rewards as a result of the outcome of the task, which relate to incentives and reward contingencies; and c) internalized extrinsic motivations, such as contributing to solve a problem of personal use benefit, which are somewhere in between two extremes of internal and external motivations.

Several theoretical studies, rooted in behavioral science and social psychology, have suggested that social factors such as reputation, image, and identification/recognition have a substantial effect on developers’ contribution to OSS communities [37, 39]. However, very few large-scale empirical studies have been conducted to verify or quantify the effect of these factors on contribution. Moreover, the results of these studies are mixed. For example, Grewal et al.’s [24] results suggest a significant relationship between developers’ network centrality and their contribution in their constructed network. However, the more recent study of Singh [46] does not find any significant effect confirming the previous results.

This study employs a unique longitudinal data set containing information of over 4 million OSS developers and their social interactions across 7 years on the popular OSS platform GitHub (www.github.com). We empirically investigate the effect of online social interaction on OSS developers' contribution levels. Our empirical result shows that receiving more followers has a positive effect on the developers' contribution level. Specifically, receiving one more follower in the previous month increases the developers' contribution level by 43.55% on average in the current month, according to our fixed effect panel data model. The effect even carries over to the next month although the marginal effect decreases. The behavior of following other developers, however, does not seem to have any effect on the contribution level of the follower according to our panel data model. In order to account for the reverse effect of people's contribution level to the number of followers they obtain, we apply a panel vector autoregression (VAR) model to study the dynamic relationship between online following and the motivation of the members to contribute. Despite a large stream of literature examining the effect of social interactions, few studies have addressed this dynamic relationship. Our result of the panel VAR model confirms the findings in previous panel data models that obtaining more followers stimulates developers to contribute more on the platform after controlling for the reverse effect of receiving followers due to their contribution levels. We employ a negative binomial model as a robustness check of the over-dispersion issue in the data set. The regression result from the negative binomial model confirms the findings in previous models.

We then try to explore other potential mechanisms that might motivate people to contribute, for example, the popular phenomenon of "likes" on social networks. People might be motivated to contribute because they observe that other people like their projects.[1] It is also possible that there are people who have developed some popular applications before and therefore receive many issue

[1] We thank an anonymous reviewer for suggesting this idea.

reports from other users. They might consider this as a kind of social obligation to resolve those issues, and hence end up contributing more. We identified two variables that could capture these characteristics of users. The variable *Stars* keeps track of how many "likes" a developer receives for her projects on the platform. Another variable is *Issues*, which keeps track of how many issue reports a developer receives from other developers for all her projects on the platform. We find that the number of *Stars* does have positive and significant effect on people's levels of contribution. The effect is however much weaker compared to that of the number of new followers. This suggests that simply being "liked" by other users might not provide enough incentive for people to make more contributions on the platform. This makes sense since people may like many users and many projects on the GitHub platform, but they choose to follow some users only when they like them to some extent. The number of *Issues* also has positive and significant effect on people's level of contribution. This suggests that besides the motivation of obtaining social recognition from their peers' following behavior, social obligation of responding to the issue reports raised by other users also positively affects people's contribution levels on the platform.

We also find that the effect of receiving new followers on one's contribution level is stronger for freelancers than those with company affiliation. Since companies are increasingly relying on open collaboration communities such as GitHub to recruit tech talent, one's contribution level and one's number of followers on the platform is a signal of one's programming and problem-solving abilities, which are the characteristics that hiring managers and recruiters are looking for.[2] We believe that there are career-related incentives besides the non-monetary incentives that previous literature has identified behind people's willingness to contribute freely on open collaboration platforms. These career-related incentives refer to potential job opportunities with monetary rewards in the future, for

[2] See http://www.cnet.com/news/forget-linkedin-companies-turn-to-github-to-find-tech-talent/, and http://www.socialtalent.co/blog/how-to-use-github-to-find-super-talented-developers (last accessed: March 8, 2018).

people making free contributions to platforms like GitHub. Our empirical results are consistent with the theoretical conclusions of Lerner and Tirole [33]. These findings have implications for researchers and provide guidelines for contributors in open collaboration communities.

Literature Review and Theoretical Background

Our study draws upon three streams of research, which range from the broader context of social behavior in general to the specific context of contribution to OSS communities: 1) image motivation and prosocial behavior, 2) social factors and contribution in online communities, and 3) social factors and contribution in OSS communities. The following sections explore each of these streams in relation to our study.

Image Motivation and Prosocial Behavior

Image or a person's desire for social approval plays an important role in how people behave in public settings. For example, since prosocial behavior signals a positive image, people are more likely to act prosaically in the public sphere (Ariely, et al. [8]). Several experimental studies, including Andreoni and Petrie [3], Dana et al [17], and Soetevent [47], have established that image is an important part of the motivation in prosocial behavior, both in controlled lab settings and in field experiments. These classical economic studies have largely been conducted in offline settings. However, more recent studies have suggested that the image effect might also hold true in online settings where people's actions are publicly observable through various online mechanisms. While the nature of both social approval and prosocial behavior differ in such settings, the image motivation of actors arguably creates similar behavioral incentives. For example, in online social media platforms, social approval might be expressed through the act of "following" and the most common prosocial behavior within online contexts is contributing to the platform by posting new content. One of the recent studies that empirically investigate such effects in online settings is by Toubia and Stephen [50]. The authors conducted an interesting field experiment on Twitter to examine the effect of social or image-related factors on members' contribution or posting ("tweeting") content on the platform. They created 100 fake Twitter accounts and gradually added followers to a group of real members over a 50-day period. They observed that although the level of activity was increased in the treated group, the effect was not statistically significant.

While the broader context of our study is the relation between people's social image and their prosocial behavior, we specifically focus our attention to online settings. Following previous studies, we consider the acts of following and liking ("starring") in online settings as proxies for social interactions enhancing a recipient's social image; similarly, a user's contribution is considered a proxy for prosocial behavior in online settings. In the following section, we briefly review recent studies on the effect of social interactions in online communities on users' contributions, in different online (electronic) communities such as organizational electronic communities [31, 51], online/virtual communities [37, 55, 56], user generated content [12, 23, 48, 50], and collaborative content communities such as open content [39]. Figure 1 summarizes these online communities.

[Please Insert Figure 1 Here.]

Social Factors and Contribution in Online Communities

Intra-organizational and inter-organizational electronic communities have been of particular interest to organizations as a means of communication, knowledge sharing, and knowledge acquisition for employees. Electronic (online) networks of practice is an example of such inter-organizational communities that provide a forum where people with a shared interest voluntarily exchange ideas and solutions to common problems. Similarly, electronic knowledge repositories (EKR) provide internal stores for knowledge sharing, project reviews, case studies, lessons learned, and best practices among the organization's employees. Several empirical studies have been conducted to examine potential factors, including social factors, affecting members' contributions to these platforms. For example, Wasko and Faraj [51, 52] conducted multiple surveys on computer professional and legal professional online networks of practice to answer why individuals help strangers in these electronic networks. The surveys' results suggest that people contribute to online networks of practice more often and with a higher quality when they perceive that it enhances their professional reputations and standing or status. Similarly, Kankanhalli et al. [31] surveyed 150 knowledge management executives, covering seven industries, to find important factors affecting knowledge contribution to organizational knowledge repositories. Their findings, in contrast to Wasko and Faraj's [51, 52] results, suggest that reputation is not a significant indicator of contribution to the knowledge repositories.

Aside from their importance to organizations, user-generated contents are fundamental to online and virtual communities in the context of social communities. As online communities are becoming ubiquitous and increasingly relevant to business, it is important for both ecommerce managers and MIS scholars to understand members' motivation to contribute to online communities ([11], [32], and [58]). Several recent studies identify the motivations to share and contribute within online communities. For example, Zhang and Zhu [55] investigate the causal relationship between group size and incentives to contribute to the Chinese-language Wikipedia through an exogenous reduction in group size as a result of the blocking of the website by the Chinese government. They found that contribution level of contributors who are not blocked shrinks by 42.8% on average, the cause of which has been attributed to loss of social benefits because of the blocking. In another study, Zhang and Wang [56] find that the contribution level and the effort of Wikipedia editors are influenced by their position in the collaboration network as well. Qiu et al. [41] examine how learning and network effects drive the diffusion of online videos on YouTube. They found that both mechanisms have statistically and economically significant effects on video views. Susarla et al. [48] consider the role of word-of-mouth (WOM) communications structured through a network and examine how cascades of WOM interactions enhance the popularity of videos on YouTube.

Ma and Agarwal [37] surveyed 650 members of two different online communities (an emotional support community and a sport-car-owners community) to find factors influencing and facilitating knowledge sharing. Similarly, Chiu et al. [15] surveyed 310 members of a virtual community of computer experts to identify the motivation underlying individuals' contribution to such communities. All these studies support the hypothesis that social factors such as reputation significantly and positively affect quantity of knowledge sharing.

Online communities are not always focused on sharing technical knowledge and expertise or providing emotional support. Online product review communities are becoming increasingly popular and social. Such websites create value by presenting an aggregated opinion of consumers on the products they review. Gathering a higher number of reviews is therefore of great interest to such platforms. Many websites encourage social interactions among users in order to solicit more reviews. Goes et al. [23] empirically study the contributions and social interactions of 92,094 members of a product review website. They observe that receiving more incoming ties increases the number of product review articles that a user posts but with a decreasing rate. Similarly, Chen et al.'s [13] natural field experiment involving 398 members on a movie review website shows that revealing social information such as social comparison leads to 530 percent increase in the number of monthly movie ratings in a group of users with a low level of activity. Interestingly, they observe a 62 percent decrease in monthly rating for a group of users with a high level of activity, at the same time.

Social Factors and Contribution in OSS Communities

The focus of this study is on OSS communities, a type of collaborative content communities where digital content is created by a group of individuals working together. In our classification of electronic communities, we place OSS communities next to open content communities (see Figure 1). Open contents are defined as creative work that others can copy or modify [53]. The contents in both types of communities are created by voluntarily collaboration of knowledgeable/skilled individuals and the changes are managed by version control systems. Several empirical studies (refer to Okoli and Oh [39]) have demonstrated how social ties have a significant effect on users' contribution to Wikipedia. However, few large-scale empirical studies have studied the effect of social factors on contribution in OSS communities. Moreover, the results of these studies are mixed. For example, Grewal et al. [24] find a significant relationship between developers' network centrality and their contribution while Singh [46] does not find any significant effect. We believe that the lack of an empirical agreement in such studies arises for two reasons. First, the social factor measures in such studies are usually approximated by other measures. For example, in the two afore-mentioned studies, two developers are assumed to be socially connected if they both have contributed to the same project at any point in time. Although in the absence of true social factors, using such proxies is the only viable option, they are not always accurate measures of those constructs. Second, OSS projects are extremely diverse: software projects range from mobile applications to desktop software to online games. Similar to commercial software packages, OSS are designed and developed for different purposes and require different sets of skills and resources. Therefore, a relationship or pattern identified in one study based on one sample of homogenous OSS projects may not hold in other settings.

Both the limitations of approximation and sampling are primarily caused by the lack of data. In this work, we seek to investigate if the afore-mentioned effects exist in OSS communities using our unique large-scale dataset of OSS community members' contributions and social activities. In other words, we seek to answer whether OSS development has become a social activity, and if so, to what extent.

Hypothesis Development

As mentioned in the previous section, previous studies have suggested that popularity has a positive effect on contribution in online communities (e.g. Ma and Agarwal [37] and Goes et al. [23]). While the studied communities have been mostly centered around knowledge sharing and user generated contents, we believe the effect might exist in the context of online community of open source development, since contribution to software projects could be considered as a specific type of knowledge contribution.

The motivation of people to engage in prosocial behavior on online platforms partially derives from their desire for social approval or acceptance. This quest for recognition from other members on the platform provides some utility that can be linked to an increase of self-image for the average user. We expect to observe an increase in a developer's contribution when she obtains more followers. Each act of "following" could be an indicator of a peer developer's interest in the work of the developer who is being "followed." Obtaining more "followers" over time might increase the developer's reputation or recognition on the platform.

H1a: Receiving more followers positively affects developers' contribution level

Note that this effect is temporal and the act of "following" happens before the act of contribution. Therefore, H1a translates to "*Receiving more followers at time t-1 positively affects developers' contribution level at time t.*"

As the number of followers increases, however, the marginal effect of obtaining a new follower should decrease since this increase is less noticeable (e.g. an increase from 92 to 93 followers is less noticeable compared to an increase from 2 to 3 followers). Therefore, the decreasing marginal effect is linked to the perception of the level of increase in the number of followers by the followed person. This diminishing marginal effect is consistent with Baumeister and Leary [9]'s argument: "the formation of further social attachments beyond that minimal level should be subject to diminishing returns; that is, people should experience less satisfaction on formation of such extra relationships," (p. 500). In view of this, it is reasonable to hypothesize the following:

H1b: The positive effect of new followers on contribution is marginally diminishing

In their seminal work, Lerner and Tirole [33] argue that by working on open source projects, programmers could improve their performance in the mission endowed by their employer, which provides them with an intrinsic utility. In other cases, the open source projects are "cool" and more fun to work on than routine tasks, which also brings some immediate benefits to them. However, the first step in contributing to a project is to be aware of the project. Theories of online communities stablished that "people begin participating as peripheral readers, proceeding to contribute to the community only after observing the community or consuming the content it has created" (Kane and Ransbotham [30]). In the OSS context, and specifically on GitHub, following a user creates a channel to directly observe that user's current and future projects and activities, and this observation might lead to the observer's increased awareness and potential contribution. Kane and Ransbotham [30] hypothesize that these theories might be a reason behind an increase in contribution to Wikipedia following an increase in consumption of the content (observation). Additionally, following and observing other active users might enhance the network (social) contagion effect observed in large social networks (e.g. Aral and Nicolaides [5], Christakis and Fowler [16]) and result in a higher level of activity. Building upon this literature, we propose that users' following behavior (consumption) on OSS platform indicates their interest and willingness to learn and to work on certain OSS projects (contribution).

H2a: Following more users positively affects a developer's contribution level

While following more members increases the chance of exposure to interesting projects, this increase is not linear in terms of the number of followed members for two reasons: 1) There is a limit in the number of projects a person can get involved with at any time and therefore awareness about more projects does not always result in contribution to more projects, and 2) the followed members might work on common projects and following more members does not always result in awareness channels to new projects.

Therefore, we assume this marginal effect of following on the level of contribution diminishes due to gaining a marginally decreasing utility for reasons similar to those explained in H1b. Thus, we have the following hypothesis:

H2b: The positive effect of following users of contribution is marginally diminishing

In regards to the effect of new followers on contribution for freelancers versus employed users, two contradicting working hypotheses could be proposed: 1) based on image motivation, and 2) based on career concern.

1) Following the argument in our Hypothesis H1a, receiving new followers increases the contribution level by leveraging image motivation and the desire for social approval. Additionally, members who spend time to complete their public profile and provide extra information about themselves to the community are more likely to be concerned about the community's perception of

them. Thus, one might expect to see a higher effect of these people' image motivation on their contribution behavior. Note that the relation between profile completeness and contribution might be merely a correlation, meaning people who have an intention to contribute will also bother to present a more detailed profile.

2) On the other hand, we have witnessed the industry taking advice from Lerner and Tirole's [33] study in recent years. The sense of selfless participation in open-source projects is becoming increasingly attractive to hiring companies. In the global market for highly-skilled software engineers, hiring managers and recruiters are realizing that the traditional hiring process which consider the candidates' work history rather than their actual work are less effective in spotting quality talent. Consequently, they are increasingly turning from LinkedIn to GitHub, where they could find not just the potential employees who look good on paper, but the ones that publicly demonstrate their capabilities on the platform.[3] Developers who maintain a good GitHub profile nowadays have an advantage over those who do not. Their GitHub profile has become the new resume for these knowledge workers.[4] In addition, potential hiring managers and recruiters might be among the new followers of a freelancer, which provides additional incentive for freelancers to demonstrate their programing ability by contributing more actively. In view of this, we believe that people who are not affiliated with a company might be more motivated to contribute by new followers. Both arguments are theoretically plausible, which presents a viable opportunity for our empirical test. We propose our hypothesis in agreement with the second argument as follows.

[3] http://www.cnet.com/news/forget-linkedin-companies-turn-to-github-to-find-tech-talent/ (last accessed: March 8, 2018).

[4] http://www.socialtalent.co/blog/how-to-use-github-to-find-super-talented-developers (last accessed: March 8, 2018).

H3. The effect of receiving new followers to the contribution level of people without announced affiliations is stronger.

Research Setting and Data

Research Context

GitHub uses a distributed version control system. Version control, or revision control, is a system that records changes over time so that any specific state (version) can be recalled later. In the simplest case, version control is analogous to Wikipedia's "page history" that enables everyone to track each change to the page. GitHub uses the "fork and pull" model for collaboration. A developer may "fork" an existing OSS project to create a personal copy of the project without requiring access to be granted to the project. Any change or "commit" to the project is first performed on this personal copy. The developer may then send a "pull" request to the maintainer of the project to merge the changes into the source project. For example, as shown in Figure 2 (the screenshot is current as of May 2016), the developers who contribute to the Linux Kernel project hosted on GitHub have submitted 601, 683 commits to the project (https://github.com/torvalds/linux). These commits include adding, modifying, or deleting pieces of code or explanatory comments. Since version control requires all information about each contribution to each project to be recorded, the detailed information about each commit is publicly available on the platform.

[Please Insert Figure 2 Here.]

In addition to developers' contributions, GitHub collects all information about developers' interactions such as "following" another developer. As shown in Figure 3, for each developer on the platform, a profile page exists which presents three types of information: 1) general information such as the developer's name, location, company and personal website (left side of the page); 2) information about each public activity that the developer performs, including her contributions to the projects, from the time she registers on the website (middle part of the page); and 3) the developer's social interactions including total number of "followers," "following," and "stars" received from other developers (lower left of the page). The detailed information on each social interaction with timestamps can be retrieved using GitHub's application program interface (API). This host of information provides a valuable resource for research on OSS.

[Please Insert Figure 3 Here.]

Data Source

GitHub is currently the largest open source community and code host in the world, with more than 4 million software developers working on more than 9 million OSS projects. Desktop software, web development, mobile applications, games, and operating systems are among the various types of OSS projects hosted on this platform. These projects are developed using various programming languages such as Java, C#, Python, PhP, JavaScript, and Ruby. Many well-known OSS project such as Linux, Bootstrap, JQuery, and Firefox for iOS are mainly or partially hosted on GitHub. For each OSS project, not only all the source codes are publicly available, but also information on every change to the project, every issue reported and/or solved, and every discussion around the project is recorded with their timestamps.

In addition, for each developer, three sets of information are recorded: 1) general information such as developer's location, company and personal website; 2) information on every public activity that a developer performs, including his contributions to the projects; and 3) the developer's social interactions including "following" and being "followed" by other developers. The exact time for each user's development activity or social interaction is recorded and stored.

Recently, a large-scale dataset on GitHub became available to researchers. This dataset spans over a various set of OSS projects and settings and contains information on the direct social interactions of the developers on the platform. It also keeps track of the developers' contributions to the platform. The dataset is well maintained and publicly accessible. All these recorded information is now publicly available, which provides an invaluable resource for researchers. In particular, what makes this dataset ideal for the purpose of this study is that the system records the exact time each "tie" (i.e. action of following a developer) has been created between developers. Using this piece of information, we were able to replicate the dynamic social network of the developers from the time the first developer joined the website (in October 2007) until the end of our study. An archive of all public activities on GitHub can be found at https://www.githubarchive.org/. A few other researchers have recently employed part of this data in their empirical analyses ([54, 57]).

The level of social interactions on the platform sharply declined (by around 75%) after December 2013. This large decrease was likely to be caused by a structural change in the design of the platform. To avoid estimation bias due to this shock, the study period is chosen to include all the activities on the platform from March 2012 to December 2013. This is also a time frame where there's no missing contribution data of the developers. We aggregate all user' interactions and contributions in each month, creating 22 periods. The choice of month as an aggregation period is consistent with research involving this dataset [57]. During this period, there were 3,266,949 registered users on the platform. Over half (1,801,674) of them, however, were inactive users. During our study period, these users made no contribution to the platform at all; further, they did not follow, nor were followed by other users. These users are irrelevant to our research. We therefore exclude these users, leaving us with those active users who either made at least one commit on the platform, or interacted with (followed or were followed by) at least one user during our study period.

Among these remaining active users there are 302,540 users who made no contributions to the platform and the only recorded behavior of them is to follow other users or get followed by others. Furthermore, there are also 628,614 users who made commits to the platform without any social interactions: they did not follow other users nor did other users follow them. Since our main interest lies in studying the effect of "following" on users' contribution levels, we retain both of these two groups of users in our dataset. This left with us a total number of 1,465,275 active users, out of which we randomly selected 10,000 for our sample. The following section describes the main variables used in our analysis.

Data Description

The notation and symbols used in the study are listed below.

- $i \in$ 10,000 users registered on the GitHub platform
- $t \in$ Months, calendar month, from March 2012 to December 2013
- $Tenure_i =$ Number of days user i has registered on the platform till March 2012
- $CompanyBinary_i =$

  Dummy variable indicating whether user i is affiliated with a company or not
- $TotalFollowers_i =$ Number of followers of user i by March 2012
- $TotalFollowing_i =$ Number of users followed by user i by March 2012
- $NewFollowers_{it} =$ Number of new followers of user i in month t
- $NewFollowing_{it} =$ Number of new users followed by user i in month t
- $Commit_{it} =$ Number of commits made by user i in month t

- $CommitSelf_{it}$ = Number of commits made by user i in month t to her own projects
- $CommitOthers_{it}$ =

  Number of commits made by user i in month t to other people's projects

Table 1 shows the summary statistics of the main variables which will be used in our analysis. The data consists of 10,000 users spanning across 22 months from March 2012 to December 2013. The first variable *Tenure* shows that the average time these users had registered on the platform is 1715 days or around four and a half years at the time our study begins. During this time, on average they had been followed by 1.55 users and had followed 1.3 users. Further, these users made on average 1.86 commits, followed 0.08 users per month and were followed by 0.09 users each month during the study period.

[Please Insert Table 1 Here.]

Empirical Model and Results

Developers on OSS contribute to projects in different ways. Writing a piece of code, revising a code written by other developers, reporting an issue in a project, and resolving a reported issue are common types of contribution. On GitHub, a contribution to a project is called a "commit". Consistent with previous literature on OSS development [24, 46], we measure a developer's performance with the number of commits in a specific time period.

Aggregating developers' activities over month, we construct a panel dataset where each unit of observation is a developer and each time period is one calendar month. We include the user's number of commits as the dependent variable and the number of individuals whom the developer follows, and the number of individuals who follow the developer during each month as the main independent variables. In addition, for each developer, the total number of her followers, the total number of the people followed by her, the number of days she had been in the system, and a binary variable for affiliation with a company are included in the model as control variables.[5]

Panel Data Methods

We first employ Random Effect and Fixed Effect regression models (see Equation 1) on the data. Subsequently, Hausman test shows that the unobserved individual effect is correlated with the regressors and the preferred model is fixed effect. The regression results in Table 2 show a statistically significant positive effect of the number of new followers (incoming ties) on users' level of contributions. On average, obtaining one more follower in last month corresponds to an increase in contribution by 0.81 commits in current month according to our fixed effect model, an increase of 43.55% compared to the average contribution level 1.86. Moreover, this positive effect stays over time. It stimulates the user to contribute 0.43 more commits in the month after, which is an increase of 23.12% compared to the average contribution level. The negative and significant coefficient of the squared of lagged *NewFollowers* suggests this effect is marginally decreasing as we have hypothesized.

$$\begin{aligned} Commit_{it} = {} & \alpha_0 + \alpha_1 NewFollowers_{i(t-1)} + \alpha_2 NewFollowers_{i(t-2)} \\ & + \alpha_3 NewFollowing_{i(t-1)} + \alpha_4 NewFollowers_{i(t-1)}^2 \\ & + \alpha_5 NewFollowers_{i(t-2)}^2 + \alpha_6 NewFollowing_{i(t-1)}^2 \\ & + \alpha_7 TotalFollowers_i + \alpha_8 TotalFollowing_i + \alpha_9 Tenure_i \\ & + \alpha_{10} Company_i + u_i + \epsilon_{it} \end{aligned} \tag{1}$$

In our third model, we include the lags of dependent variable as covariates and run a dynamic panel data estimation. By applying the within (demeaning) transformation or by taking first differences, the fixed effect model helps us deal with unobserved heterogeneity, such as the developers' creativity or ability which might be directly correlated with their contribution to the platform. However, as

[5] As the regression result shows below, these time-invariant variables will be cancelled out in the fixed effect model.

Nickell [38] shows, the demeaning process of fixed effect model that subtracts the individual's mean value of dependent and each independent variable from the respective variable creates a correlation between regressor and error, making the standard estimators inconsistent. In our third model, we adopt a Dynamic Panel Data (DPD) approach to address this problem. The DPD approach was suggested by Arellano and Bond [7], who popularized the work of Holtz-Eakin, Newey and Rosen [28]. They derived a consistent generalized method of moments (GMM) estimator and specified the model as a system of equations, one per time period using all available lags of the untransformed variables as instruments. The lag of dependent variable is included as one of the regressors to adjust for serial correlation. The estimator is designed for datasets with many panels (large N) and few periods (small T) and thus is appropriate for our data in providing unbiased estimates. Equation 2 specifies our empirical estimation using Arellano Bond model.

$$\begin{aligned} Commit_{it} = \beta_0 &+ \beta_1 Commit_{i(t-1)} + \beta_2 NewFollowers_{i(t-1)} \\ &+ \beta_3 NewFollowers_{i(t-2)} + \beta_4 NewFollowing_{i(t-1)} \\ &+ \beta_5 NewFollowers_{i(t-1)}^2 + \beta_6 NewFollowers_{i(t-2)}^2 \\ &+ \beta_7 NewFollowing_{i(t-1)}^2 + u_i + \epsilon_{it} \end{aligned} \tag{2}$$

The regression result of Model 3 in Table 2 confirms the result of our random effect and fixed effect models and verifies Hypothesis 1 that receiving more followers in previous month does increase users' level of contribution in current month and the marginal effect is decreasing.

[Please Insert Table 2 Here.]

The coefficient of lagged *NewFollowing* is not significant for all three models, which suggests that the number of individuals followed by a developer does not have a significant effect on her contribution level. This is consistent with our expectations: there are many users, such as novice programmers, who registered an account on the platform just to learn from other more experienced developers by following them and watching their projects without necessarily contributing to the platform.

A Panel VAR Model

In the previous section, we showed that obtaining more followers could positively affect the users' contribution level to the online community. On social platforms like GitHub, the typical user could easily be attracted to heavy users who make a large number of contributions to the platform on prominent projects. This indicates that people's contribution level on the platform could very well influence others to follow them. In this section, to account for this reverse effect of people's contribution level on their number of followers and further study the dynamic relationship between online following and online members' motivation to contribute, we employ a panel VAR [1] approach to identify the Granger causality, while controlling for individual heterogeneity and time effects. Panel VAR model is a type of dynamic panel model that assumes that each dependent variable is a function of its own past values and the past values of all other dependent variables. The strength of panel VAR models is that it allows us to control for unobserved individual heterogeneity and utilize lagged dependent variables as instruments within the model in the GMM estimation to obtain consistent estimates. We find that other people's following behavior has a significant positive effect on users' contribution level to the platform even after controlling for the aforementioned factors. On the other hand, users' contribution level also has a significant positive effect on number of followers they receive. These results have implications for online community designers and OSS scholars.

In the prior literature, VAR is widely used in Macroeconomics (e.g. [22]), and recently it has been increasingly adopted in the IS area [12, 20, 35]. The main advantages of our panel VAR approach is that (a) it allows us to examine the bi-directional relationships between online following and online members' motivation to contribute through Granger causality tests; and (b) the dynamics between online following and online members' motivation to contribute over time can be assessed and visualized through techniques such as impulse response functions.

The specification of our model has the following form:

$$\mathbf{Y}_{i,t} = A(L)\mathbf{Y}_{i,t} + \boldsymbol{\beta} \cdot \boldsymbol{\Gamma}_{i,t} + \boldsymbol{\delta}_i + \boldsymbol{\theta}_t + \varepsilon_{i,t},$$

where $\mathbf{Y}_{i,t}$ is a vector of covariates as follows:

$$\mathbf{Y}_{i,t} = \begin{bmatrix} Commit_{i,t} \\ NewFollowers_{i,t} \\ NewFollowing_{i,t} \end{bmatrix},$$

$\boldsymbol{\Gamma}_{i,t}$ is a vector of time-varying control variables, $\boldsymbol{\delta}_i$ is the individual fixed effects characterizing the members' time-invariant attributes, and $\boldsymbol{\theta}_t$ is set of time dummies that control for any time effects such as seasonality. The lag operator L is defined by $L\mathbf{Y}_{i,t} = \mathbf{Y}_{i,t-1}$, and we also define the symbol $L^p\mathbf{Y}_{i,t} = \mathbf{Y}_{i,t-p}$. Let $A(L)$ be the lag polynomial:

$$A(L) = a_1 L + \cdots + a_p L^p,$$

which is defined as an operator such that

$$A(L)\mathbf{Y}_{i,t} = a_1 \cdot \mathbf{Y}_{i,t-1} + \cdots + a_p \cdot \mathbf{Y}_{i,t-p}.$$

In our panel VAR model, we use the Helmert transformation to control for individual fixed effects. This forward differencing procedure overcomes the problem that fixed effects and lagged dependent variables are inherently correlated [6, 22, 34].

Before estimating the model, we consider the model selection issue and calculate the model selection measures for first to third-order panel VARs using the first four lags of *Commit*, *NewFollowers*, and *NewFollowing* as instruments. The result is shown in Table 3.

[Please Insert Table 3 Here.]

Based on the three model selection criteria by Andrews and Lu [4], we should pick the model with the smallest Bayesian Information Criterion (MBIC), Akaike Information Criterion (MAIC) and

Hannan-Quinn Information Criterion (MQIC). In general, the first-order panel VAR is the preferred model because it has the smallest MBIC and MQIC. Based on the selection criteria, we estimate the first-order panel VAR model. We also improve the estimation efficiency by using "GMM-style" instruments as proposed by Holtz-Eakin, Newey, and Rosen [28].

The estimation results for our panel VAR model are shown in Table 4. Our main objective is to examine whether the number of followers of the members has a significant effect on the latter's contributions to the platform after controlling for other factors such as individual heterogeneity and time effects. Since the number of lags picked according to the above model selection criteria is 1, this approach of one-period lagged dependent variables allows us to directly interpret the short-term effect of each independent variable on the dependent variable. It is worthwhile to point out that the interpretation of the coefficient estimates on dependent variables with lagged periods of more than one is not straightforward since the effect depends not only on one particular variable but also on other lagged dependent variables. Usually, impulse response functions (IRFs) are used to visually interpret the coefficient estimates in these cases. We show them at the end of this section.

From Table 4 we can see that, in the *Commit* equation, the coefficient estimate on first lag of *NewFollowers* is positive and significant at the 1% level, indicating that obtaining new followers in previous month has a positive effect on users' contribution level of current month. From the *NewFollowers* equation, the positive and significant coefficient estimate on *Commit* at lag 1 indicates that the reverse effect indeed exists, i.e., people's contribution level of previous month has a positive effect on number of followers they receive in current month. We also notice in the *Commit* equation, the coefficient estimate on *NewFollowing* at lag 1 is negative, which is not found in our random effect and fixed effect model.

[Please Insert Table 4 Here.]

We perform the Granger causality test and the results are shown in Table 5: *NewFollowers* Granger-causes *Commit* at the 1% confidence level, conversely, *Commit* also Granger-causes *NewFollowersers* at the 5% confidence level. These test results indicate that number of new followers obtained by a user in the past helps predict her contribution level in the future and vice versa. The test also shows that people's number of commits in the past possesses predictive value for number of users they will follow in the future. The intuition could be that a high level of contribution indicates that the user is highly motivated. The motivation could come from contributing to her own project or from contributing to other people's projects. In order to do the latter, she will surf on the platform, look for projects she is interested in, follow the project owner and make commits to those projects.

[Please Insert Table 5 Here.]

We also check the stability condition of the estimated panel VAR by calculating the modulus of each eigenvalue of the estimated model. A VAR model is stable if all moduli of the companion matrix are strictly less than one [26, 36]. The eigenvalues shown in Table 6 and Figure 4 show that all the eigenvalues lie inside the unit circle, and thus, pVAR satisfies stability condition [27].

[Please Insert Table 6 Here.]

[Please Insert Figure 4 Here.]

We supplement our regression estimates with the analysis of the corresponding impulse response functions (IRFs). The IRFs provide us with an intuitive way to observe the response of one variable to a shock in another variable and inspect whether the impact is perennial or short-lived. The IRF confidence intervals are computed using 200 Monte Carlo draws based on the estimated model. Figures 5 and 6 highlight selective IRFs so we can examine the response of users' level of contribution to a shock in number of followers they gain and vice versa. From Figure 5, we see that the reaction of *Commit* to a shock in *NewFollowers* is positive and it stays positive over time until after 10 periods (i.e. 10 months) it decreases to close to zero. Additionally, we see the significant effect of *Commit* on *NewFollowers* in Figure 6. There is an immediate reaction of *NewFollowers* to a shock in *Commit* and the reaction stays positive and attenuates slowly over time.

[Please Insert Figure 5 Here.]

[Please Insert Figure 6 Here.]

Robustness Checks

Log-transformed Regression

Among the over one million active users in our dataset, there are many who made very few or even no commits to the platform, did not follow anyone, nor was followed by anyone. There are, on the other hand, also a few extremely active and popular users who made a large number of commits, followed many users, and were followed by many users. Therefore, the distribution of the count variables is highly skewed. In this section, we perform log-transformed random and fixed-effect panel-data model as a robustness check. The regression result is shown in Table 7 in online appendix. The result is consistent with our previous analysis. The number of followers people received from last month has a positive and significant effect on their contribution level of the current month. The effect also carries over to the next month. Holding other predictor variables fixed, we expect about a 2.7% increase in people's contribution level during current month when their number of followers increases by 10% in last month, which further leads to 2.3% increase of contribution in the next month. The negative coefficient on the squared term indicates that the marginal effect is decreasing.

Over-dispersion and Other Possible Explanatory Variables

So far, we have discussed the effect of following on people's contribution level on the open source platform. Other than this mechanism, there might be other factors motivating people to contribute:

for example, the popular phenomenon of "likes" on social networks. People might be motivated to contribute because other people like their projects. It is also highly possible that some people receive many issue reports from other users. Such reports might create a kind of social obligation to resolve those issues, which might lead these users to contribute more. Within our dataset, we identified two variables that could capture these characteristics of users. The variable *Stars* keeps track of how many "likes" a developer receives for her projects on the platform. Another variable is *Issues*, which keeps track of how many issue reports a developer receives from other developers for all her projects on the platform. Table 8 in online appendix shows the correlation matrix of the main variables. The strongest correlation coefficient is 0.8543 which is between *Commit* and *CommitSelf*. The next one is 0.5534 which is between *NewFollowers* and Stars. The correlation between *NewFollowers* and *Issues* is not strong, with the coefficient being 0.185. We include these two new variables, *Stars* and *Issues*, into our model and run the following random effect and fixed effect panel data model. The results are in Table 9 in online appendix.

$$\begin{aligned} Commit_{it} = {} & \alpha_0 + \alpha_1 NewFollowers_{i(t-1)} + \alpha_2 NewFollowers_{i(t-2)} \\ & + \alpha_3 NewFollowing_{i(t-1)} + \alpha_4 Stars_{i(t-1)} + \alpha_5 Issues_{i(t-1)} \\ & + \alpha_6 NewFollowers_{i(t-1)}^2 + \alpha_7 NewFollowers_{i(t-2)}^2 \\ & + \alpha_8 NewFollowing_{i(t-1)}^2 + \alpha_9 Stars_{i(t-1)}^2 + \alpha_{10} Issues_{i(t-1)}^2 \\ & + \alpha_{11} TotalFollowers_i + \alpha_{12} TotalFollowing_i + \alpha_{13} Tenure_i \\ & + \alpha_{14} Company_i + u_i + \epsilon_{it} \end{aligned} \tag{3}$$

Table 9 (in online appendix) shows that after counting for Stars and Issues, receiving new followers in previous month still has positive and significant effect on people's contribution level of current month and the effect carries over to the next month. According to the fixed effect model, one more follower from last month increases the number of commits by 0.772 after controlling for Stars and Issues, which is not significantly different from the magnitude in the model not controlling for these two variables. We also see that number of *Stars* does have a positive and significant effect on the developers' level of contribution; however, the effect is much weaker compared to that of number of new followers. This suggests that simply being liked by other users might not provide enough incentive for people to make more contributions on the platform; what motivates them to contribute more is being followed by others. Developers on GitHub may superficially like many users and many projects on the platform, but they choose to follow only some users with a greater level of commitment. The number of *Issues* also has positive and significant effect on people's level of contribution. This suggests that besides the motivation of obtaining social recognition from their peers' following behavior, social obligation of responding to the issue reports raised by other users also positively affects people's contribution level on the platform.

From Table 1 (Summary Statistics of Main Variables), we can see that the variance of the count variables is much larger than the mean, which suggests for over-dispersion in the count variables. To address this issue and examine the robustness of our previous models, we employ the negative binomial model as robustness check. Table 10 in online appendix shows the regression result of negative binomial model that verifies the results of our previous models. The random-effects, fixed-effects and population-averaged negative binomial models all suggest that receiving more followers in previous month has positive and significant effect on users' contribution of current month and the effect carries over to the next month.

Company Affiliation and Larger Sample Analysis

On each user's profile page, there is a field describing whether this user is affiliated with a company or not. In our dataset we use the dummy variable *CompanyBinary* to keep track of this. Table 11 in online appendix summarizes the statistics of key variables for these two groups of users from our over one-million-active users. It shows that the contribution level of people with company affiliation are more active than freelancers. During each month of our study period, on average, they have more new followers, follow more users and contribute more, both to their own and to other users' projects.

To investigate the difference in the effect of following behavior to the contribution level between these two groups of users, we include into our model an interaction term *CompanyBinaryL1NewFollowers* by interacting the company affiliation dummy variable with lag of number of new followers. To further check the robustness of our model, we perform a larger random sampling of 200,000 users out of the 1,465,275 active users. We then employ the following random effect and fixed effect panel data model. The regression results are shown in Table 12 in online appendix.

$$\begin{aligned} Commit_{it} = \alpha_0 &+ \alpha_1 NewFollowers_{i(t-1)} + \alpha_2 NewFollowers_{i(t-2)} \\ &+ \alpha_3 NewFollowing_{i(t-1)} + \alpha_4 CompanyBinaryL1NewFollowers \\ &+ \alpha_5 NewFollowers_{i(t-1)}^2 + \alpha_6 NewFollowers_{i(t-2)}^2 \\ &+ \alpha_7 NewFollowing_{i(t-1)}^2 + \alpha_8 TotalFollowers_i \\ &+ \alpha_9 TotalFollowing_i + \alpha_{10} Tenure_i + \alpha_{11} Company_i + u_i + \epsilon_{it} \end{aligned} \quad (4)$$

The negative coefficient of the interaction term suggests that the effect of followers to a user's contribution level is much stronger for freelancers than those who are affiliated with a company. This implies that freelancers care more and get more motivated by the number of followers they obtain than those who are affiliated with a company. A possible reason is that people's contribution level and number of followers on the platform is a signal of their programing and problem solving ability. As we mentioned earlier, companies are turning from LinkedIn to GitHub to find tech talent and a user's GitHub profile is becoming the new resume for job hunters. Moreover, the number of followers is one of the most important information recruiters specifically look at when they go through users' profiles. This provides motivation for those freelancers who are hunting for jobs to show their ability and their popularity to potential employers by getting more followers on the open source platform.

Limitation and Future Research Directions

While we identified the effect of following on developers' contribution levels through a unique data set on GitHub, we recognize several limitations of our work. First, panel VAR model is just a first step, which helps us establish the Granger causality between developers' online following behavior and their level of contribution. In order to establish the causal impact of following on individuals' contributions more convincingly, we need natural or field experiments to further confirm the results. Second, in our company affiliation analysis, it is possible that some users are affiliated with a company but did not include this information in their profile, which would diminish the difference between the two groups and make our estimation for the affiliated group biased downward.

We propose two enhancements, for future research, that improve our models and analyses. First, enriching the current analyses by including more factors, such as the number of projects in which a developer is involved, in the model. Second, surveying a set of developers using their publicly available email addresses and including the information in the model.

Managerial Implications

Out results and findings have important implications for both the stakeholder groups in OSS development process: the OSS platforms as well as the OSS community.

Implications for OSS Platforms: While GitHub is the leading platform in OSS development, there has always been many other OSS platforms (e.g. SourgeForge, BitBucket) that vary in terms of features, designs, and communities. Two main categories of features on such platforms are related to software development (e.g. creating, managing and "commiting" to software) and social interaction (following, liking, reporting issues). As platform owners continuously improve their platforms by updating or adding features, they have to decide how to allocate their resources on each feature category. To make an informed decision, the platform owners and designers need to know the importance of each feature to the overall success of their platform. Our research confirms and quantifies the effect of social factors on users' contribution level, which is an important success measure for the OSS platforms. These findings will inform platform owners about the importance of the social factors and help them make better decisions on the features that can maximize their impact. For example, to the best of our knowledge, no OSS platform has features to facilitate finding and following users using matching methods such as recommender systems. Our results suggest that it might be worthwhile for platform developers to invest on such efforts to increase social interactions between users.

In addition, our results suggest that freelancers are highly motivated by receiving new followers. Currently, GitHub does not have any mechanism to notify the followed member, besides updating the number of followers on her profile. We believe directly notifying a member when another member follows her might be an effective strategy to incentivize higher contribution, particularly for freelancers.

Implications for OSS Community: It is important for the members of OSS community to understand the value of each form of social interaction on the OSS platforms. Showing appreciation to OSS developers for their work is a common practice on such platforms and is often explicitly encouraged.[6] For example, GitHub states that starring a developer's project performs the action of "[S]howing appreciation to the repository maintainer for their work." Often OSS members would like to see more progress in a developer's project that they are using and our research indicates that showing appreciation encourages more work on the project. However, our results suggest that

[6] https://help.github.com/articles/about-stars/

following a developer has significantly higher impact on a developer's contribution level than merely liking (starring) the project. Therefore, we recommend the OSS community to consider the use of the "following" feature more prominently on OSS platforms to incentivize higher contribution levels to their projects.

Conclusion

Online communities have been growing rapidly during the last two decades in various forms, from user-generated content and social media to collaborative-content communities. While actively contributing on these platforms, people connect to each other and make new friendships. Open Source Software (OSS) development platform is one of these communities. The extant literature has tried to find the incentives for people to contribute to the creation of public goods such as OSS, even when they are not paid for their efforts. Although several theoretical studies have suggested that motivations in the form of reputation and recognition have substantial effect on people's contribution to OSS communities (see [42] and theoretical studies on motivation cited by [44]), very few large-scale empirical studies have been conducted to verify or quantify the effect of these factors. Some existing results of these studies are conflicting, and the social interactions have usually been approximated by other measures [24, 45]. Using a unique longitudinal dataset within an OSS environment that contains information of over 4 million users and their social interactions across 7 years, we empirically investigate the effect of online social interactions on people's contribution levels.

We find that recognition from other users in the previous month in the form of following has a significant positive effect on people's contribution levels to the community in the current month even after controlling for individual heterogeneity. The effect even carries over to the next month although the marginal effect decreases. This finding is consistent with the conclusion of previous studies [23, 52] that audience size has a significant effect on people's contribution level in such online open communities.

We also find that the number of "likes" has a positive and significant effect on people's levels of contribution, although the effect is much weaker compared to that of the number of new followers. This suggests that simply being liked by other users might not provide enough incentive for people to make more contributions on the platform. This is what one might expect, since a user may like many other users and many projects on the platform, but she chooses to follow only users with whom she shares stronger ties (e.g. shared projects). The number of *Issues*, such as bug reports, also has positive and significant effect on people's level of contribution. This suggests that besides the motivation of obtaining social recognition from their peers' following behavior, social obligation of responding to the issue reports raised by other users also positively affects people's contribution level on the platform.

We further observe that the effect of being followed on a user's contribution level is much stronger for freelancers than those who are affiliated with a company. Nowadays, many IT companies are turning from LinkedIn to GitHub to find the technical talent that they need and are using the users' GitHub profile as reference. Our findings suggest that besides the non-monetary incentives that previous literature has found that behind people's free contribution to open collaboration platforms, there are career-related incentives too: contribution to the OSS platforms can lead to future monetary rewards. Our empirical results are consistent with the theoretical conclusions of Lerner and Tirole [33]. We did not find a consistent significant effect for the number of developers followed by a user on her own contribution level across the different statistical models that we employed. We believe that simply following other developers is not an indication of intent to contribute to the community, since there are many novice programmers who follow others in order to learn.

Appendix: Figures and Tables

Figure 1. Different electronic communities and members' motivations to contribute

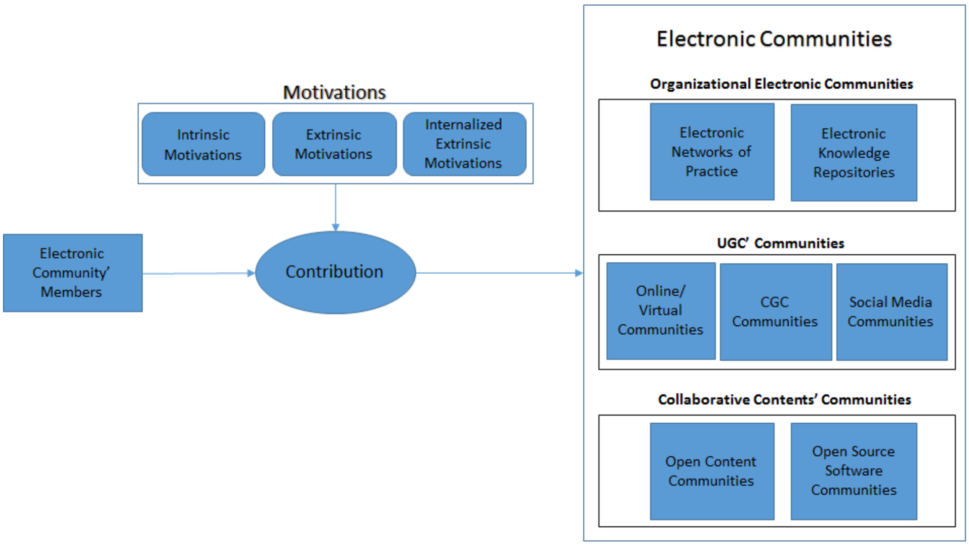


Figure 2. Linux, an OSS project on GitHub

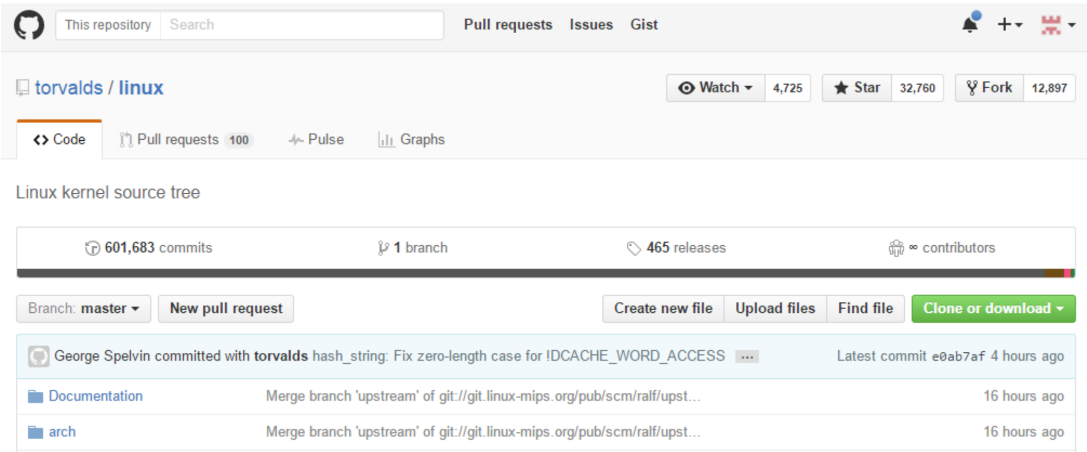


Figure 3. A user's profile on GitHub including public activities and social factors

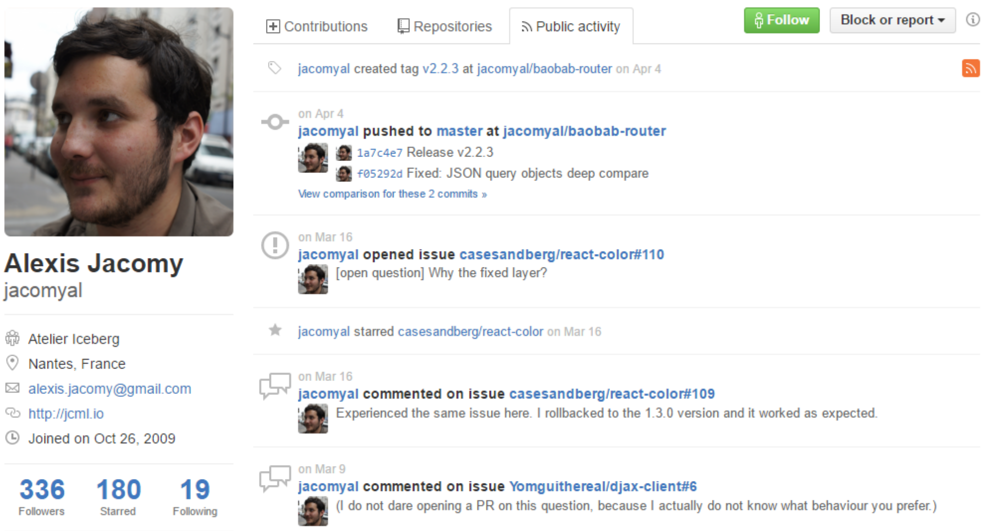


Table 1. Summary statistics of main variables

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Tenure	220,000	1715.49	445.90	122	2283
CompanyBinary	220,000	0.12	0.33	0	1
TotalFollowers	220,000	1.55	32.25	0	2972
TotalFollowing	220,000	1.30	8.75	0	634
NewFollowers	220,000	0.09	1.26	0	204
NewFollowing	220,000	0.08	1.01	0	151
Commit	220,000	1.86	18.34	0	3850
CommitSelf	220,000	1.34	15.28	0	3850
*Commit*Others	220,000	0.53	9.54	0	2196
Stars	220,000	0.29	8.25	0	1852
Issues	220,000	0.06	1.62	0	338

Table 2. Random effect, fixed effect and Arellano Bond models

VARIABLES	(1) Random Effect	(2) Fixed Effect	(3) Arellano Bond
L.NewFollowers	1.102***	0.811***	1.159***
	(0.237)	(0.207)	(0.0633)
L2.NewFollowers	0.733***	0.430***	0.826***
	(0.174)	(0.149)	(0.0603)
L.NewFollowing	0.127	0.0871	-0.00559
	(0.148)	(0.161)	(0.0532)
Lag1NewFollowersSq	-0.00593***	-0.00413***	-0.00116***
	(0.00129)	(0.000931)	(0.000398)
Lag2NewFollowersSq	-0.00486***	-0.00289***	-0.00121***
	(0.00111)	(0.000737)	(0.000390)
Lag1NewFollowingSq	-0.00148	-0.00110	-0.000217
	(0.00111)	(0.00119)	(0.000529)
TotalFollowers	0.0255		
	(0.0165)		
TotalFollowing	0.0875		
	(0.0666)		
Tenure	-0.00184***		
	(0.000338)		
CompanyBinary	0.763**		
	(0.314)		
L.Commit			-0.301***
			(0.00228)

Note: Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 3. Panel VAR model selection

lag	CD	J	J pvalue	MBIC	MAIC	MQIC
1	.8060909	57.76852	.0005133	-267.4074	3.768525	-76.61207
2	.8363832	29.80703	.0393654	-186.9769	-6.192973	-59.78004
3	.8236389	19.65672	.0201532	-88.73526	1.656722	-25.13681

Table 4. Panel VAR estimation for total contribution

	Dependent variable		
Independent variable	*Commit*	*NewFollowers*	*NewFollowing*
L.*Commit*	0.685***	0.00256**	0.00178***
	(0.199)	(0.00129)	(0.000532)
L.*NewFollowers*	4.112***	0.489***	0.0187***
	(1.342)	(0.0880)	(0.00613)
L.*NewFollowing*	-2.222***	-0.0743***	0.0492***
	(0.841)	(0.0271)	(0.0132)

Robust Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5. Panel VAR-Granger Causality Wald Test

Equation \ Excluded		chi2	Df	Prob > chi2
Commit	*NewFollowers*	9.391	1	0.002
	NewFollowing	6.981	1	0.008
	All	12.592	2	0.002
NewFollowers	*Commit*	3.924	1	0.048
	NewFollowing	7.531	1	0.006
	All	11.661	2	0.003
NewFollowing	*Commit*	11.184	1	0.001
	NewFollowers	9.333	1	0.002
	All	21.098	2	0.000

Ho: Excluded variable does not Granger-cause Equation variable
Ha: Excluded variable Granger-causes Equation variable

Table 6. Eigenvalue stability condition

Eigenvalue		
Real	Imaginary	Modulus
0.718204	0	0.7203034
0.2523988	0	0.4465782
0.0406613	0	0.056595

Figure 4. Eigenvalues lie inside the unit circle

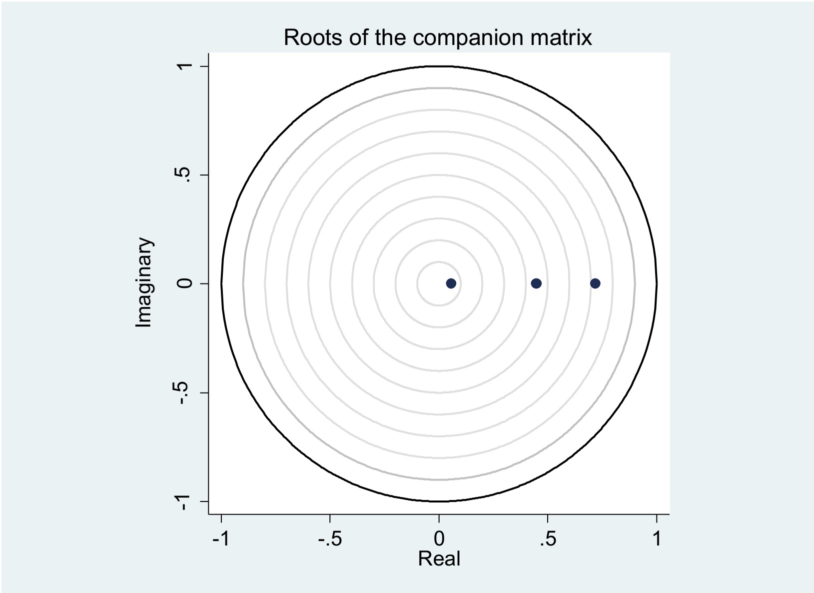


Figure 5. Response of C*ommit* to *NewFollowers*

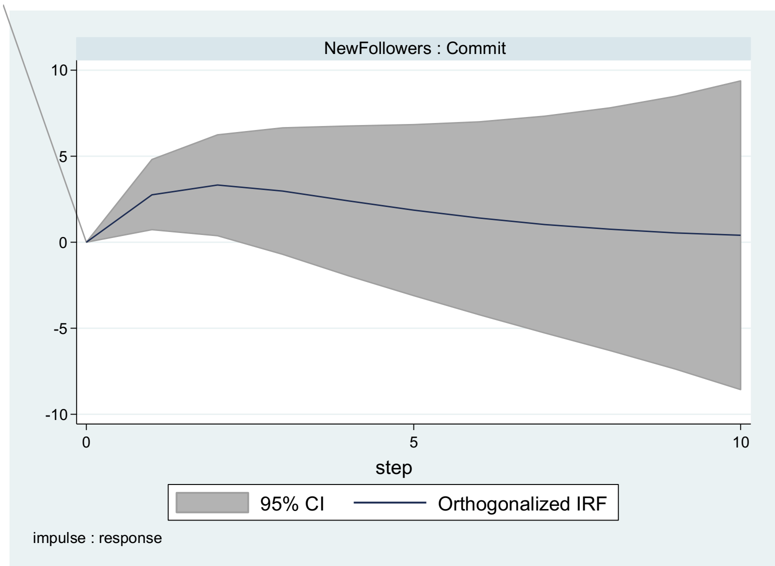
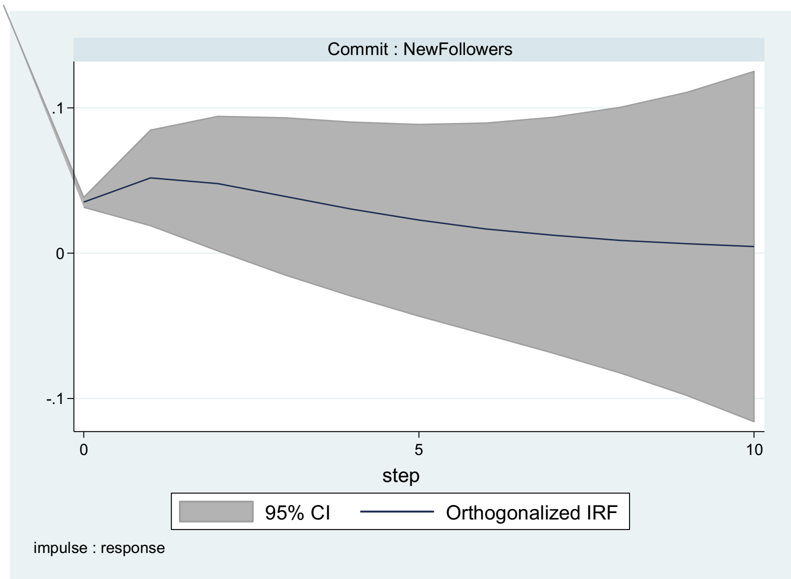


Figure 6. Response of *NewFollowers* to *Commit*



References

1. Abrigo, M.R. and Love I. *Estimation of Panel Vector Autoregression in Stata: A Package of Programs* (2015), manuscript available on http://paneldataconference2015.ceu.hu/Program/Michael-Abrigo.pdf.
2. Aksulu, A. and Wade M.A. comprehensive review and synthesis of open source research. *Journal of the Association for Information Systems,* 11, 11 (2010), 576-656.
3. Andreoni, J. and Petrie, R. Public goods experiments without confidentiality: A Glimpse into Fund-Raising. *Journal of Public Economics*, 88, 7-8 (2004), 1605-23.
4. Andrews, D.W. and Lu, B. Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics*, 101, 1 (2001), 123-164.
5. Aral, S, and Nicolaides, C. Exercise contagion in a global social network. *Nature* communications 8 (2017), 14753.
6. Arellano, M. and Bond S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58, 2 (1991), 277-297.
7. Arellano, M. and Bover, O. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68, 1 (1995), 29-51.
8. Ariely, D., Bracha, A., and Meier, S. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, 99, 1 (2009), 544-55
9. Baumeister, R.F. and Leary, M.R. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 3 (1995), 497.
10. Bonaccorsi, A. and Rossi, C. Comparing motivations of individual programmers and firms to take part in the open source movement: From community to business. *Knowledge, Technology & Policy, 1*8, 4 (2006), 40–64.
11. Bulgurcu, B., Van Osch, W., and Kane, G.C. The rise of the promoters: user classes and contribution patterns in enterprise social media. *Journal of Management Information Systems*, 35, 2 (2018), 610-646.
12. Chen, H., De, P., and Hu, Y.J. IT-enabled broadcasting in social media: An empirical study of artists' activities and music sales. *Information Systems Research*, 26, 3 (2015), 513-531.
13. Chen, Y., Harper, F.M., Konstan, J., and Xin Li, S. Social comparisons and contributions to online communities: A field experiment on movielens. *The American Economic Review*, 100, 4 (2010), 1358-1398.
14. Cheng, H.K., Liu, Y., and Tang, Q. The impact of network externalities on the competition between open source and proprietary software. *Journal of Management Information Systems*, 27, 4 (2011), 201-230.
15. Chiu, C., Hsu, M., and Wang, E.T. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. *Decision Support Systems* 42, 3 (2006), 1872–1888.
16. Christakis N.A. and Fowler J.H. The collective dynamics of smoking in a large social network. *New England journal of medicine* 358, 21 (2008), 2249-58.
17. Crowston, K., Wei, K., Howison, J. and Wiggins, A. Free/Libre open-source software development: What we know and what we do not know. *ACM Computing Surveys* (CSUR) 44, 2 (2012), 7.
18. Dana, J., Cain, D.M. and Dawes, R.M. What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games. *Organizational Behavior and Human Decision Processes*, 100, 2 (2006), 193-201

19. DeWall, C.N., Baumeister, R.F. and Vohs, K.D. Satiated with belongingness? Effects of acceptance, rejection, and task framing on self-regulatory performance. *Journal of Personality and Social Psychology*, 95, 6 (2008), 1367.
20. Dewan, S., and Ramaprasad, J. Social media, traditional media, and music sales. *MIS Quarterly*, 38, 1 (2014), 101-121.
21. Fang, Y., & Neufeld, D. Understanding sustained participation in open source software projects. *Journal of Management Information Systems*, 25, 4 (2009), 9-50.
22. Fort, T.C., Haltiwanger, J., Jarmin, R.S., and Miranda, J. How firms respond to business cycles: The role of firm age and firm size. *IMF Economic Review*, 61, 3 (2013), 520-559.
23. Goes, Paulo B., Lin, M. and Yeung, Ching-man Au. Popularity effect in user-generated content: evidence from online product reviews. *Information Systems Research,* 25, 2 (2014), 222–238.
24. Grewal, R., Lilien, Gary L. and Mallapragada, G. Location, location, location: how network embeddedness affects project success in open source systems. *Management Science*, 52, 7 (2006), 1043–1056.
25. Grossman, L. *New Free License to Cover Content Online*. Netly News, 1998.
26. Hamilton, J.D. *Time series analysis* (Vol. 2). Princeton: Princeton university press, 1994.
27. Harris, R.D., and Tzavalis, E. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics*, 91, 2 (1999), 201-226.
28. Holtz-Eakin, D., Newey, W., and Rosen, H. Estimating vector autoregressions with panel data. *Econometrica*, 56, 6 (1988), 1371-95.
29. Höst, M. and Oručević-Alagić, A. A systematic review of research on open source software in commercial software product development. *Information and Software Technology,* 53, 6 (2011), 616-624.
30. Kane, G.C., and Ransbotham, S. Content as community regulator: The recursive relationship between consumption and contribution in open collaboration communities. *Organization Science*, (2016).
31. Kankanhalli, A., Tan, Bernard CY, and Wei, Kwok-Kee. Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, March 1 (2005), 113–143.
32. Khansa, L., Ma, X., Liginlal, D., and Kim, S.S. Understanding members' active participation in online question-and-answer communities: A theory and empirical analysis. *Journal of Management Information Systems*, 32, 2 (2015), 162-203.
33. Lerner, J. and Tirole, J. Some simple economics of open source. *The Journal of Industrial Economics,* 50, 2 (2002), 197–234.
34. Love, I., and Zicchino, L. Financial development and dynamic investment behavior: Evidence from panel VAR. *Quarterly Review of Economics and Finance*, 46, 2 (2006), 190-210.
35. Luo, X., Zhang, J., and Duan, W. Social media and firm equity value. *Information Systems Research*, 24, 1 (2013), 146-163.
36. Lütkepohl, H. New Introduction to Multiple Time Series Analysis. *Springer Science & Business Media*, 2005.
37. Ma, M. and Agarwal, R. Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research*, 18, 1 (2007), 42–67.
38. Nickell, S. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, (1981), 1417-1426.
39. Okoli, C. and Oh, W. Investigating recognition-based performance in an open content community: A social capital perspective. *Information & Management,* 44, no. 3 (2007), 240–52.
40. Osterloh, M and Rota, S. Open source software development—Just another case of collective

invention? *Research Policy*, 36, 2 (2007), 157–1571.
41. Qiu, L., Tang, Q., and Whinston, A.B. Two Formulas for Success in Social Media: Learning and Network Effects. *Journal of Management Information Systems*, 32, 4 (2015), 78-108.
42. Roberts, J.A., Hann, I. 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 (2006), 984–999.
43. Sacks, M. Competition between open source and proprietary software: Strategies for survival. *Journal of Management Information Systems*, 32, 3 (2015), 268-295.
44. Sen, R.A. strategic analysis of competition between open source and proprietary software. *Journal of Management Information Systems*, 24, 1 (2007), 233-257
45. Singh, P.V. The small-world effect: The influence of macro-level properties of developer collaboration networks on open-source project success. *ACM Transactions on Software Engineering and Methodology (TOSEM),* 20, 2 (2010), 6.
46. Singh, P.V., Tan, Y. and Mookerjee, V. Network effects: The influence of structural social capital on open source project success. *MIS Quarterly*, 35, 4 (2011), 813-829.
47. Soetevent, A.R. Anonymity in Giving in a Natural Context - an Economic Field Experiment in 30 Churches. *Journal of Public Economics*, 89, 11-12 (2005), 2301-23.
48. Susarla, A., Oh, J.H., and Tan, Y. Influentials, Imitables, or Susceptibles? Virality and Word-of-Mouth Conversations in Online Social Networks. *Journal of Management Information Systems*, 33, 1 (2016), 139-170
49. Temizkan, O., and Kumar, R.L. Exploitation and Exploration Networks in Open Source Software Development: An Artifact-Level Analysis. *Journal of Management Information Systems*, 32, 1 (2015), 116-150
50. Toubia, O. and Stephen, Andrew T. Intrinsic vs. image-related utility in social media: Why do people contribute content to twitter? *Marketing Science,* 32, 3 (2013), 368-392.
51. Wasko, M. and Faraj, S. It is what one does: Why people participate and help others in electronic communities of practice. *The Journal of Strategic Information Systems* 9, 2 (2000), 155–73.
52. Wasko, M. and Faraj, S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, (2005), 35–57.
53. Wiley, D. Open Content, OpenContent.org. (19Online] Available at http://www.opencontent.org/definition/
54. Zampetti, F., Ponzanelli, L., Bavota, G., Mocci, A., Di Penta, M. and Lanza, M., How developers document pull requests with external references. *IEEE/ACM 25th International Conference* (2017), 23-33
55. Zhang, X. and Zhu, F. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *The American Economic Review*, 101, 4 (2011), 1601-1615.
56. Zhang, X., and Wang, C. Network positions and contributions to online public goods: The case of Chinese Wikipedia. *Journal of Management Information Systems*, 29, 2 (2012), 11-40.
57. Zhang, Y., Yin, G., Yu, Y. and Wang, H., Investigating social media in GitHub's pull-requests: a case study on Ruby on Rails. *Proceedings of the 1st International Workshop on Crowd-based Software Development Methods and Technologies* (2014), 37-41
58. Zhao, L., Detlor, B., and Connelly, C.E. Sharing knowledge in social Q&A sites: the unintended consequences of extrinsic motivation. *Journal of Management Information Systems*, 33, 1 (2016), 70-100.