The Influence of Project Initiators' Person-to-Person Followership on Project Popularity in Open Source Communities: The Role of Reach and Importance

Qin Weng

Frank Soh

Abstract

Person-to-person (P2P) followership is an important aspect of major open source software (OSS) development platforms. In this age of social media platforms, P2P followership has significantly shaped the way users engage in OSS development by facilitating the establishment of connections among OSS users. Despite the prevalence of P2P followership, less is known about the impact of OSS project initiators' P2P followership nodes on their project popularity. This is a particularly important gap considering the low rate of OSS project success. We posit that OSS project initiators derive information and influence benefits from the quantity and connectivity of their P2P followership nodes, explaining the popularity of the projects initiated. We determine the connectivity of OSS project initiators' P2P followership nodes based on the nodes' reach and importance. To test the hypotheses, we use a large panel dataset collected over 24 months from GitHub. The findings indicate that the quantity of OSS project initiators' P2P followership nodes including followers and followees has a positive effect on their project popularity. Moreover, such an effect is mostly dependent on the connectivity of the OSS project initiators' P2P followership nodes in such a way that highly connected P2P followership nodes do not impact influence benefits but they increase information benefits. We discuss the theoretical and practical implications of our findings.

Keywords

open source software project; project initiator; followership; following; social network; social media

Highlights

- Person-to-person (P2P) followership affects OSS initiators' project popularity
- OSS initiators gain information benefits from their followee networks
- OSS initiators gain influence benefits from their follower networks
- Large GitHub panel datasets are used to test the hypotheses
- P2P followership positively affects initiators' project popularity; highly connected P2P followership nodes increase information benefits

The Influence of Project Initiators' Person-to-Person Followership on Project Popularity in Open Source Communities: The Role of Reach and Importance

Introduction

Open source software (OSS) development offers a significant market expected to reach $47.50 billion by 2025 (MarketResearchEngine 2021). Individual developers and organizations – including tech leaders such as Microsoft, Facebook, and Amazon – are increasingly relying on OSS development platforms (e.g., GitHub) to create and distribute software (e.g., Microsoft has more than 4,000 repositories on GitHub as of July 2021). Unlike traditional (i.e., close source) software development, OSS development provides access to an unlimited pool of developers located worldwide. OSS development is becoming an essential part of software firms' strategy (Morgan and Finnegan 2014), contributing to the overall success of these firms (Gulati et al. 2012). Unfortunately, despite the growing importance of OSS development (e.g., Eclipse, Linux, and MySQL) as a credible alternative to traditional software development – about 50% of Fortune 10 firms use OSS development (Daniel et al. 2018) – only a small number of OSS projects achieve success (Chengalur-Smith and Sidorova 2003; Lin et al. 2017).

OSS research has greatly enhanced our understanding of OSS success factors (e.g., Setia et al. 2012). However, the landscape of OSS platforms is continuously evolving. For instance, in addition to traditional virtual team setups, OSS platforms may offer *social media capabilities* so that members of the community can follow each other. This is the case with GitHub wherein users can follow each other, thus increasing the users' number of followers and followees (see Figure 1). However, on traditional OSS platforms, such a feature is mostly missing. These social media afforded capabilities challenge our traditional wisdom on the antecedents and outcomes of OSS development efforts. They further suggest the potential of a people-centric approach rather than the traditional project-centric approach to studying OSS success. "The mutual constitution of technologies and online communities' sociality is still understudied" (Faraj and Shimizu 2018, p. 12). These OSS development platforms, also known as *social coding platforms*, are hypercompetitive environments. For example, GitHub has over 29 million project initiators, 70 million registered users, and 170 million repositories as of May 2021 (Github 2021). Given the wide range of projects from which OSS users can choose, OSS project initiators face challenges in increasing their projects' popularity. For example, about 95% of GitHub repositories had no stars as of January 2021. On OSS development platforms, the most popular projects tend to appear on the front pages of the website. With the large abundance of projects in GitHub today, such prestige becomes ever more important. Also, popular projects may attract attention from businesses and recruiters. Thus, popularity can be a powerful incentive and reward for contributors as it opens up more and better career opportunities.

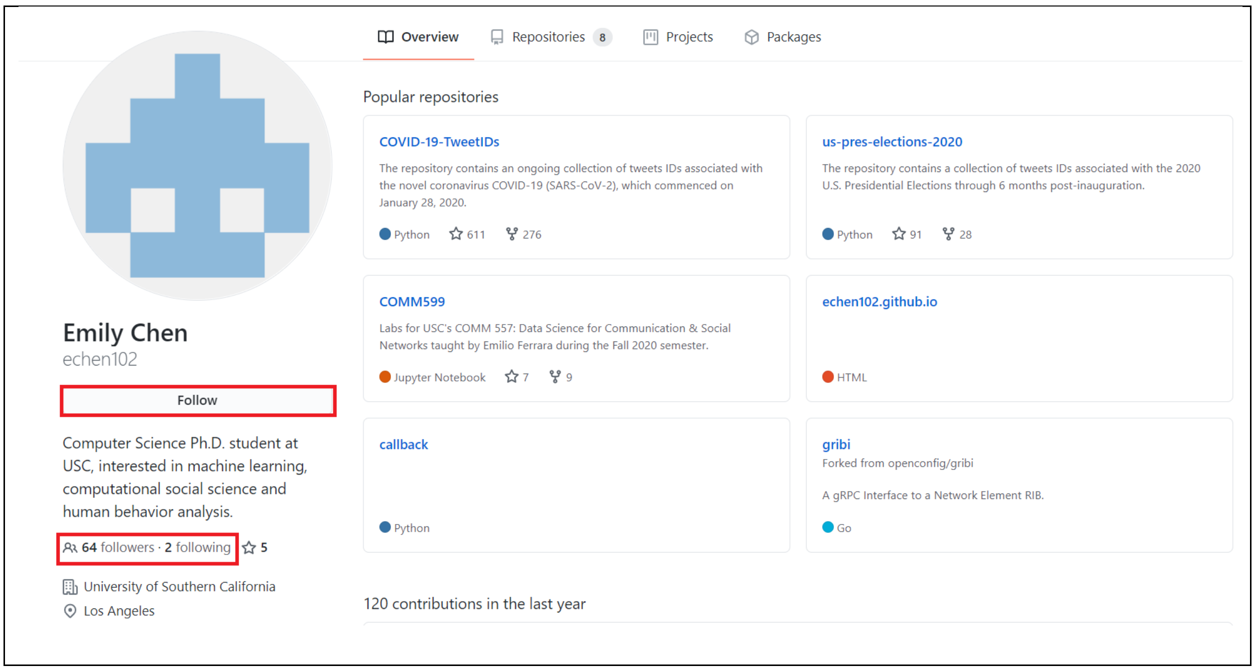


Figure 1. Numbers of Followers and Followees of a User on GitHub

Understanding the factors that influence project popularity has strategic importance for OSS project initiators. Yet, limited OSS research incorporates the aforementioned social media capabilities to identify the potential antecedents of an initiator's project popularity. Scholars have repeatedly used social network theory to understand how OSS projects' and developers' networks influence OSS project popularity (Peng et al. 2013; Sutanto et al. 2021). Despite the great potential of the theory, the types of networks focused on in OSS literature are still limited and do not fully capture the unique social media afforded capabilities of such OSS platforms as GitHub. Specifically, our review of the literature indicates that most studies on OSS project popularity focus on *affiliation networks* (i.e., projects are connected if they have common contributors) and *user* *support network*s (i.e., users are connected if they communicate) (Peng et al. 2013; Sutanto et al. 2021). While these networks are important, they are not the only networks that exist in OSS development.

Recently, some OSS researchers have started to incorporate social media capabilities in their studies and examine the *followership network*; however, the burgeoning perspective is still quite limited and deserves more work. Specifically, a limited number of recent studies shed light on two additional types of networks including follower (i.e., individuals who follow a focal actor) and followee (i.e., individuals that the focal actor follows) networks, formed through the OSS platform's social media capability followership. For example, Jiang et al. (2019) and Moqri et al. (2018) highlight two mechanisms, namely learning and social recognition, through which the follower network influences code reuse and developer's contributions, respectively. In this study, we extend their work by looking at a different outcome variable and argue that follower and followee networks benefit an OSS initiator's project popularity. By doing so, we not only expand the nomological network that connects followership and success on social coding platforms, but also stay consistent with prior literature that indicates the impact of social interactions of employees on the employees' work performance in organizational settings (Sykes and Venkatesh 2017).

We refer to follower and followee networks as person-to-person (P2P) followership networks to differentiate them from affiliation networks used in prior literature. We posit that OSS project popularity is accomplished through the information and influence benefits afforded by the P2P followership (i.e., follower and followee networks) on OSS development platforms. We focus on the potential P2P followership's effect on popularity for two main reasons. First, P2P followership is prevalent across major OSS development platforms (e.g., GitHub). It is a social media capability afforded by an OSS development platform through which platform users can follow (i.e., followee network) and be followed by others (i.e., follower network). It enables the establishment of connections among platform users facilitating the sharing of users' activities (e.g., contribution activities) and the monitoring of what others are doing on the OSS development platform. Second, despite the prevalence of P2P followership connections among platform users in general (about 5.5 million registered users on GitHub have at least one follower) and OSS project initiators in particular (about 4.6 million OSS project initiators on GitHub have at least one follower), our understanding of the impact of

P2P followership networks on an OSS initiator's project popularity is still limited. Our study intends to fill the research gap by answering the following research question: *how do OSS project initiators' P2P followership networks impact their project popularity on OSS development platforms?*

Considering that both building follower networks and following others consume time and effort, understanding the effect of network effects can help OSS initiators to strategically decide two things: 1) whether to expand the followership network; and 2) what type of followers and followees are more valuable. To answer these questions, we build on the social network theory as we describe the theoretical mechanisms that explain the effects of the OSS project initiator's P2P followership networks on their project popularity. Our core thesis is that an OSS project initiator's P2P followership networks generate information and influence benefits, which in turn increase the initiator's project popularity. Importantly, we also examine the boundary conditions in which the characteristics of the network nodes affect the relationship between social networks and popularity. In the GitHub context, we test our hypotheses using a large panel dataset with 128,362 initiator-month observations across January 2012-December 2013. We thoroughly research and examine the updates and changes implemented by the GitHub platform since 2012 and confirm that there have been none that would threaten the relevance and validity of our findings using this dataset (see Appendix A for the full background of the dataset and GitHub updates). We buttress the findings with a more recent although incomplete dataset from 2016-2017 in the robustness test. The findings confirm the significance of the OSS project initiators' P2P followership nodes in explaining their project popularity.

We structure the rest of the paper into four sections. First, we present the theoretical background of the study. Next, we develop the hypotheses. We then describe the research design and report the results of the analyses. Finally, we discuss the contributions, implications, and limitations of the study.

Theoretical Background

Prior Literature on OSS Project Popularity

Project popularity is a commonly used outcome to evaluate the success of OSS projects (see Medappa and Srivastava 2019). OSS project popularity reflects the extent to which an OSS project generates interest among OSS platform users. Moreover, it represents the level of visibility of an OSS project within the

OSS community. It is particularly difficult for initiators to generate community interest in their OSS as GitHub is overcrowded with repositories. In such a context, OSS popularity is an important indicator of OSS success as OSS projects compete for developers and final users (Borges and Valente 2018). A popular project is more likely to have high sociality in terms of requests pulled, high activity in terms of commits, and high efficiency in terms of bug resolution. Thus, OSS project popularity accurately represents OSS project success.

Popularity as a success measurement has been explored in different ways in OSS studies. For example, past research has examined downloads (e.g., Daniel et al. 2013; Grewal et al. 2006; Wen et al. 2013), page views (e.g., Setia et al. 2012), and stars (e.g., Jarczyk et al. 2014; Medappa and Srivastava 2019; Tsay et al. 2014). Table 1 presents examples of outcomes that focus on OSS project popularity. Following prior studies that explore individual-level outcomes on OSS development platforms (e.g., code reuse (Jiang et al. 2019)), we shift the focus of success to the popularity of OSS projects created by their initiators. We follow prior research by focusing on stars (see Medappa and Srivastava 2019). Starring is a unique social media functionality embedded in OSS development platforms (e.g., GitHub). Starring a project is like bookmarking it, allowing a user to conveniently keep track of the project and repeatedly access it at a later time.

While there are several other OSS popularity metrics, they are not comparable. For example, GitHub relies on starring behaviors for recommendations and identifying trending OSS (Github 2017). Moreover, practitioners indicate that the number of stars is the most useful OSS popularity metric (Borges and Valente 2018). The number of stars is found to highly influence future contributions (Borges and Valente 2018; Moqri et al. 2018) and usage behaviors (Borges and Valente 2018). By starring a repository, individuals indicate their interest and appreciation in the OSS (Borges and Valente 2018; Github 2017). Besides practitioners, scholars support the usefulness of the number of stars to measure OSS popularity by using it to identify popular OSS projects (Jarczyk et al. 2014; Medappa and Srivastava 2019; Tsay et al. 2014). Hence, the number of stars constitutes a more appropriate assessment of OSS project popularity.

Table 1. Review of Prior Literature on OSS Project Popularity

Study	Predictors			Outcomes⁺
	Project Characteristics	Contributor Characteristics	Initiator Characteristics	
Stewart et al. 2006	×			Count of subscribers
Chengalur-Smith et al. 2010	×	×		Count of new developers and users
Setia et al. 2012	×	×		Count of downloads and page views
Daniel et al. 2013	×	×		Count of downloads
Peng et al. 2013	×	×		Count of downloads
Sutanto et al. 2014	×			Count of downloads and active users
Medappa and Srivastava 2019	×			Count of stars
Setia et al. 2020	×			Downloads
Sutanto et al. 2021	×			Traffic intensity
This study			×	Count of stars

[+] The list of outcomes contains only popularity-related variables.

Scholars have extensively examined factors that influence OSS project popularity. These factors can be categorized into two groups. First, project characteristics describe the attributes of OSS projects. Examples of project characteristics include project maturity (Sutanto et al. 2021), project ownership (Medappa and Srivastava 2019), and project license (Stewart et al. 2006). Second, contributor characteristics represent the attributes of project members (e.g., developers, administrators, etc.). Examples of contributor characteristics include participant's level of contribution (Daniel et al. 2013), participant's language spoken (Daniel et al. 2013), developer's position in the network (Setia et al. 2012), and participant's type (Peng et al. 2013). We present a review of prior literature on the antecedents of OSS project popularity in Table 1. Our review suggests that prior studies on OSS project popularity have overlooked the characteristics of project initiators. While other OSS studies have examined the importance of project initiators (e.g., Jiang et al. 2019), less work has attempted to investigate the significance of project initiators on OSS project popularity. This study aims at addressing this gap.

Social Network Theory

Using a social network approach, researchers can capture how people, organizations, or groups interact with others inside their network (Kilduff and Tsai 2003). Borgatti and Halgin (2011) describe a network as "a set of actors or nodes along with a set of ties of a specified type … that link them." (p. 1169). In general, there are three types of networks, including ego-centric networks, socio-centric networks, and open-system networks (Kadushin 2012, p. 17). Specifically, an ego-centric network focuses on the network of a single node; a socio-centric network focuses on networks in a closed system; an open-system network focuses on networks without clear boundaries. To represent the relationships among individual nodes, the relationship tie could be reciprocal or directional. The tie could also be strong or weak. Additionally, the structural position of an actor in a social network can be described using the concept of centrality, which includes degree, betweenness, and eigenvector (Kilduff and Tsai 2003). If we apply these concepts to the context of the OSS network, each of the individual participants becomes a node, connected by directional weak ties, which together constitute the social-centric network of the OSS community. To describe the structural position of each actor, we apply to this study the concept of degree centrality to calculate the direct ties to an actor. Additionally, we extend the concept of centrality by including the characteristics of an actor's connected nodes.

Borgatti and Foster (2003) propose two explanatory mechanisms that explain the impact of a social network. The two mechanisms differ in the way they treat ties. The first one is the connectionist approach. Here, the focus is on the flow of resources through ties. According to this mechanism, the performance of an individual in the network depends on the resources (e.g., information) that flow through their ties. The second one is the structuralist approach. Here, the focus is on the structure of ties, and not the content. This mechanism suggests that two individuals in the network have similar performance when they occupy structurally similar positions. These two mechanisms are core to social network theory.

Under the big umbrella of social network theory, some key theories include the small world theory (Travers and Milgram 1969), the strength of weak ties theory (Granovetter 1973), structural holes theory (Burt 1992), and social capital theory (Coleman 1988). These theories largely examine the importance of the positioning of an individual in the overall network, which depicts an end-state without prescribing a means how to get there. In other words, the theory focuses on the network position of the ego while ignoring the *feature* of the ego's *connected nodes*. Our goal is to fill this gap in two ways: 1) by contextualizing the network in the OSS setting; and 2) by uncovering actionable strategies by taking into account the nuanced features of an ego's connected nodes. Specifically, we contribute to the theory by highlighting the heterogeneity attributed to each follower/followee's own unique social network. We draw from social network theory and argue that social networks generate important benefits including information benefits (Granovetter 1973) and influence benefits (Coleman 1988), and that strategically following others and attracting followers help an OSS initiator to garner these benefits.

Prior OSS literature on Social Network Theory

As OSS development platforms enable P2P followership, they include a key feature of social media platforms (e.g., Twitter) which has been overlooked in prior literature on OSS project popularity. Previous studies on the impact of network structures on OSS project popularity have not highlighted P2P followership even though P2P followership connections are largely formed by OSS platform users. Researchers have instead focused on affiliation networks (Peng et al. 2013; Sutanto et al. 2021) and communication networks (Sutanto et al. 2014) (see Table 2), with little attention given to the P2P followership network.

Table 2. Review of Prior Literature on OSS Project Popularity with a Focus on Network Structure

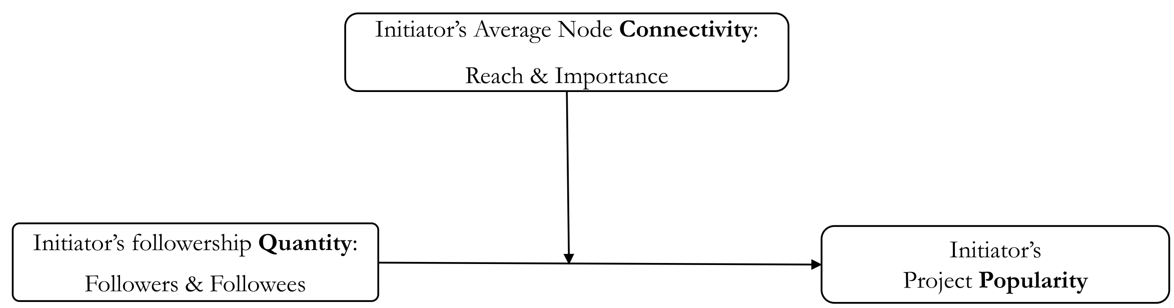
Study	**Network Type**	**Network Measures**
Peng et al. 2013	Affiliation network Node = project Link = two projects are connected if they have common contributors	Number of network ties, network tie type (leader-follower, follower-leader, etc.)
Sutanto et al. 2014	Communication network Node = OSS platform users Link = two users are connected if they are in communication	In-degree and betweenness centrality
Sutanto et al. 2021	Affiliation network Node = project Link = two projects are connected if they have common contributors	Structural hole, network cohesion
This study	P2P followership network Node = OSS platform users Link = two users are connected if one follows the second user.	Count of followers and followees

An *affiliation network* refers to a network wherein actors are connected to projects (Grewal et al. 2006), projects are connected to each other based on common contributors (Peng et al. 2013), or actors are connected to each other based on membership to common projects (Singh et al. 2011). The social ties that are formed between developers illustrate their collaboration in projects. In this approach, the ties do not represent the decision of a developer to form a tie with a non-collaborator developer. Being connected to other members of a project is not the result of a developer's choosing the connections, but a consequence of the decision to become a member of the project. A *communication network* identifies developers who are connected if they are engaged in a form of communication together. Previous research has connected people who participate in the reporting and discussion of bugs (Long 2006) and user support forums (Sutanto et al. 2014). This type of network also reflects collaboration between individuals (e.g., in the resolution of bugs).

The above two types of networks are valuable ways to perceive social connections in the OSS community, but they do not fully capture the capability of OSS development platforms to enable an OSS platform user to choose another OSS platform user – whether a collaborator or not – with whom to establish a connection. This is a missed opportunity, as P2P followership is an important capability – similar to social media capabilities – on today's OSS development platforms. While recent efforts such as Jiang et al. (2019) and Moqri et al. (2018) focus on followership connections on OSS development platforms, they do not examine OSS initiators' project popularity. Moreover, they do not highlight the connectivity-related heterogeneity across P2P followership nodes. *Connectivity* represents the connections or interactions a node has with other nodes. The emphasis of prior studies has been mostly on the quantity (i.e., number of nodes) aspect of P2P followership nodes (e.g., number of additional followers and followees (Moqri et al. 2018), and the number of non-repetitive collaborations and observers (Jiang et al. 2019)). Even though Jiang et al. (2019) distinguish between observers and developers among followers (which serves as a quality distinction), the connectivity aspect of P2P followership nodes is still overlooked, specifically, the connectivity-related heterogeneity across P2P followership nodes. Their implicit assumption is that P2P followership nodes are equal in terms of connectivity, limiting the understanding of the role of P2P followership nodes. Our study intends to fill the research gap by considering the difference of each node using its connectivity. We thus examine complementarities between the quantity and connectivity of P2P followership nodes of OSS project initiators while explaining their project popularity on OSS development platforms.

Research Model

In this study, we identify the connectivity aspect of P2P followership nodes by distinguishing each node of an OSS project initiator's P2P followership networks based on its *importance* – number of followers (i.e., individuals who follow the initiator) – and *reach* – number of followees (i.e., individuals being followed by the initiator). Moreover, we identify the quantity aspect of P2P followership nodes by counting the number of followers and followees of an OSS project initiator. We posit that the numbers of followers and the followees afford different advantages to project initiators, affecting their project popularity. We further argue that such effects depend on the average followers' connectivity and followees' connectivity in terms of their reach and importance. We present the theoretical framework in Figure 2.



1) We recognize that each follower and followee is different. We capture connectivity of each node using reach and importance.
2) We posit that the quantities of followers and followees predict an OSS initiator's project popularity, moderated by the initiator's average node connectivity.

Figure 2. Theoretical Framework

A central proposition of social network theory surrounds the resources and capabilities embedded in the networks (Kilduff and Tsai 2003). In this study, we focus on the followership observed in the OSS community, and specifically, the followership between project initiators and others in the OSS community. We posit that initiators on an OSS development platform derive influence and information benefits, which in turn affect their OSS project popularity. We use the follower network – the set of followers of an OSS project initiator – to explore the influence benefits. Moreover, we use the followee network – the set of followees of an OSS project initiator – to explore information benefits. We posit that information benefits derived from the number of followees are enhanced as the connectivity (i.e., importance and reach) of the followees of an OSS project initiator increases. In contrast to information benefits, we argue that influence benefits derived from the number of followers are reduced as the connectivity (i.e., importance and reach) of the followers of an OSS project initiator increases. Hence, our research highlights the mixed effects of the connectivity of P2P followership nodes. Figure 3 is our research model.

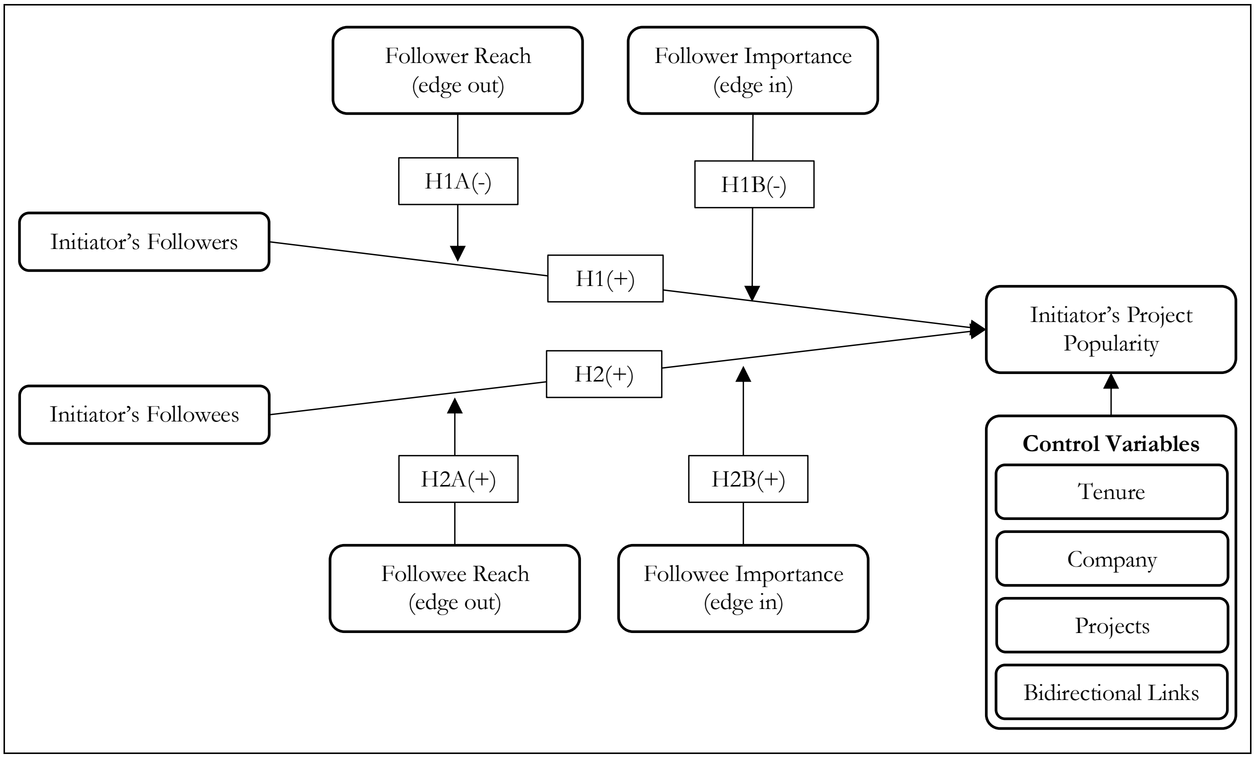
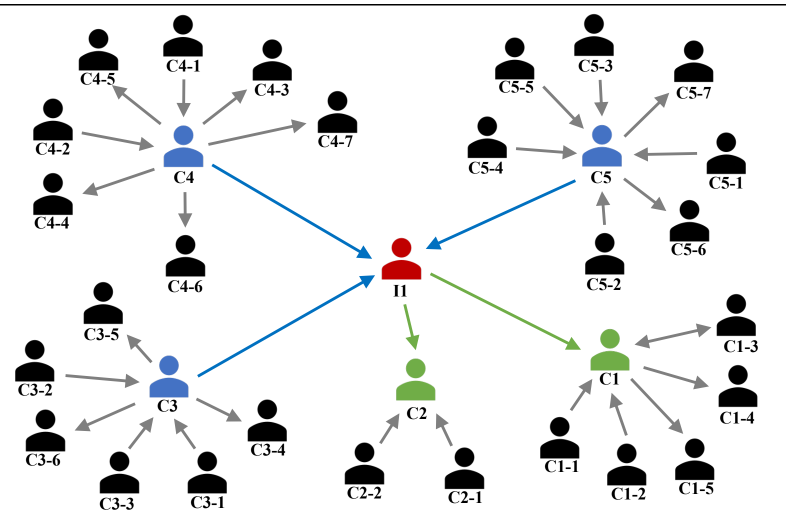


Figure 3. Research Model

P2P Followership and OSS Project Popularity

OSS development platforms increasingly enable P2P followership, suggesting that the P2P followership networks of a project initiator may influence the initiator's project popularity. OSS project initiators' P2P followership networks generate information and influence benefits, potentially leading to the initiator's project popularity. The follower network is a directed social network that represents who is following a project initiator. The followee network is a directed social network that represents whom a project initiator is following. Both the aforementioned ties are direct.

However, literature also suggests that both direct and indirect ties are important to consider (Ahuja 2000). An indirect tie occurs when two nodes are connected through other nodes in the network. We therefore further explore the characteristics of these direct ties using the indirect ties, i.e., ties of the direct ties. We use the ties of the direct followers and followees to represent two important characteristics of theirs. Specifically, in both the follower and followee networks, we examine the effects of the reach and importance of a project initiator's followers and followees. We illustrate these concepts in Figure 4. The nodes connected to the initiator (e.g., I1) with the incoming and outgoing arrows represent followers (e.g., C3, C4, C5) and followees (e.g., C1, C2), respectively. Moreover, the number of incoming arrows on either a follower or a followee represents the node importance. The number of outgoing arrows on either a follower or a followee represents the node reach. Next, we develop our hypotheses in detail. We posit that such followership affords both information benefits through the followee network and influence benefits through the follower network. These benefits in turn contribute to an OSS initiator's project popularity.



For the follower network, I1 has three followers C3, C4, C5 (shown in blue), each with a different number of followers and followees, which we refer to as connectivity related heterogeneity: importance and reach. Specifically, C5 has the highest importance as it has the most followers (=5), while C4 has the highest reach as it has the most followees (=6).

For the followee network, I1 has two followees C1 and C2 (shown in green), with C1 having higher importance (4 followers) and reach (3 followees).

Figure 4. Reach and Importance of Followers and Followees

Influence Benefits: Follower Network

The number of an initiator's followers reflects the extent of the initiator's influence benefits or the number of people the initiator can influence (Coleman 1990). It illustrates the network spread represented by the number of incoming ties linked directly to the project initiator. We consider the effect of the number of followers can be explained through both the connectionist and structuralist explanatory mechanisms. First, in the connectionist explanatory mechanism, ties are seen as conduits through which information or influence flows (Borgatti and Foster 2003). The direct notification functionality afforded by the network ensures followers get notified of project initiators' activities. Notifications raise followers' awareness of initiators' project updates and new projects, enhancing the opportunity for their projects to get starred. Thus, follower network ties are conduits facilitating the spread of information about the project initiator's activities.

Second, in the structuralist explanatory mechanism, the focus is not on the content of the ties but on the structure of the ties in the ego network (Borgatti and Foster 2003). For example, individuals in OSS development may aim to interact with influential others (Kuk 2006). The preferential attachment argument indicates that individuals form a preference for those members who are highly active and visibly well connected (Faraj and Johnson 2011). Thus, the structure of their ego network is important. For example, the number of followers enhances the quantity of contribution (Goes et al. 2014; Moqri et al. 2018). It reduces or counterbalances users' uncertainty related to project initiators and increases the trust in the initiators. The OSS community includes a large number of users coming from all over the world. Those users are more likely to be strangers to one another. Users have a high level of uncertainty about the abilities of project initiators to manage the project. They have no assurance of the expertise and professionalism of project initiators. One way to reduce such uncertainty is through looking at project initiators' network reach, which is typically transparent on OSS development platforms. Users likely assume that project initiators with a large number of followers within the community are more trustworthy in terms of expertise and abilities. All in all, the connectionist and structuralist mechanisms suggest that project initiators who attract more followers derive more influence benefits which in turn may increase the popularity of the projects they created. Therefore, we posit that:

H1: Receiving more followers positively affects the popularity of the projects initiated.

Follower reach represents the average out-degree of a project initiator's followers. Out-degree here refers to the number of people a follower follows. Follower reach is important as it relates to the influence an initiator exerts in the follower network. It reflects the strength of the tie initiator-follower. Initiators have a greater influence on individuals with whom they are tightly connected (see Krackhardt 1992). Thus, we argue that an initiator's ability to increase the popularity of the projects initiated depends on whether they have tight or loose connections (i.e., high, or low amount of community's time and attention) with followers. In online communities, tie strength is crucial as online individuals depend on the amount of the community's time and attention to thrive and succeed (Wang et al. 2013).

A follower who follows many people is likely to have loose connections with the initiator, i.e., he/she may not allocate enough time and attention to the initiator because individuals can only dedicate limited time and attention to others' activities (Ocasio 1997; Simon 1971). On OSS development platforms, anyone can easily follow another person. Therefore, the problem is not information access but information overload which in turn leads to loose connections (Shapiro and Varian 1998). There may be a limit to the number of "tight" ties that an individual can maintain (Wellman 1997). It is therefore important to consider the network reach of the followers to account for the influence a project initiator exerts on these followers. For a project initiator who attracts followers with a high network reach, i.e., people who follow many other people, the gain in popularity coming from the increased number of followers weakens due to loose connections (i.e., the less influence the project initiator exerts on the followers). Thus, we posit that:

H1A: For a project initiator whose new followers have a high network reach, the number of new followers will have a lesser positive effect on the popularity of the projects initiated.

Follower importance represents the average in-degree of a project initiator's followers. In-degree here refers to the number of followers of a follower. An initiator is likely to have loose connections with followers that have high importance because these followers tend to not give much time and attention to others. These followers are influencers themselves and connect loosely with initiators because they need to dedicate time and attention to their own followers. In other words, the strength of the tie initiator–follower is weakened because the follower's time and attention to the project initiator are distracted by their followers. Thus, a project initiator has tighter connections with followers that have lower importance. In other words, the time and attention paid to a project initiator by followers with high importance are lower than that of followers with low importance. With reduced time and attention from their followers, the influence benefits received from the follower network are likely reduced. Therefore, for a project initiator who attracts followers with high network importance, i.e., people with many followers, the gain in popularity coming from the increased number of followers weakens due to the reduced influence the project initiator exerts on the followers (i.e., loose connections). Thus, we posit that:

H1B: For a project initiator whose new followers have high network importance, the number of new followers will have a lesser positive effect on the popularity of the projects initiated.

Information Benefits: Followee Network

The effect of the number of followees can be explained through the connectionist explanatory mechanism (see Borgatti and Foster 2003). We argue that OSS project initiators with a larger number of followees receive greater information benefits in the OSS community. Network ties provide access to resources, among which a valuable one is the information benefits afforded by the network ties (Nahapiet and Ghoshal 1998). This is consistent with the "social access to resources" category identified by Borgatti and Foster (2003). According to this category (i.e., social access to resources), an individual's performance results from the *quality* and *quantity* of resources available in the network. Ties are channels through which an individual can access those resources. Project initiators who follow a large number of people within the community can get access to diverse sources of information. By actively following an individual, a project initiator not only selectively identifies the projects and sources that may be useful or of interest, but also secures access to updates and changes to the information source so that the knowledge stays current. This way the followee network influences the quality of a project initiator's work, affecting the initiator's project popularity.

An OSS community is composed of very diverse people with different backgrounds and expertise. When a project initiator maintains a large number of connections, many people may be strangers to one another offline. Such ties may be loose but effective channels for information flow, especially for searching for useful information (Hansen 1999). Individuals benefit from their loose (or weak) ties by getting access to diverse information[1], while people who are strongly connected tend to convey very similar information. Loose (or weak) ties are thus important for detecting opportunities (Granovetter 1973, 1983). A project

[1] Information overload may also be an issue here. As a result, as the number of followees increases, the marginal benefit of an additional followee may reduce.

initiator who increases the number of followees will likely increase the information benefits provided by the followee network, and in turn, produce projects that attract users' interest. Therefore, we posit that:

H2: Following more users positively affects the popularity of the projects initiated.

Followee reach represents the average out-degree of a project initiator's followees. Out-degree here refers to the number of people a followee is following. This concept is important in the followee network as it relates to the *quality* of information source for a project initiator obtained through following others. As mentioned above, in the social access to resources category, an individual's performance results from the *quality* and *quantity* of resources available in the network. The person whom the project initiator follows becomes the knowledge source for the initiator. The knowledge source (i.e., the followee) however may also follow others. People who follow many people have access to more information. Following these people is just like following people that possess broader knowledge. It functions as indirectly increasing the quantity of information access by a project initiator. Such information benefits in turn help enhance the quality of a project initiator's work as it likely reflects the broader interest of the community. For a project initiator who follows individuals with a high network reach, i.e., individuals with many followees, the gain in popularity coming from the increased number of followees increases, due to the enhanced information gain the project initiator receives from the followees. Therefore, we posit that:

H2A: For a project initiator whose new followees have a high network reach, the number of new followees will have a greater positive effect on the popularity of the projects initiated.

Followee importance represents the average in-degree of a project initiator's followees. In-degree here refers to the number of followers of a followee. Followee importance matters as it relates to the *quality* of the information an initiator obtains from the followee network. Followees who are followed by many people are important people recognized by the community. Their knowledge is representative of the interest of the OSS community. Project initiators who are following important people are behaving as if they are following people that possess expertise and knowledge. They, therefore, tend to be more aware of the community's development activities, contributions, and discussions. Such project initiators have a better understanding of the community interest and therefore can detect more opportunities in terms of project development. Based on the knowledge acquired through these high-quality sources, project initiators can create projects that attract or spark the interest of the community. For a project initiator who follows individuals with high network importance, i.e., individuals with many followers, the gain in popularity coming from the increased number of followees strengthens due to the information gain the project initiator receives from the followees. Therefore, we posit that:

H2B: For a project initiator whose new followees have high network importance, the number of new followees will have a greater positive effect on the popularity of the projects initiated.

Method

Our study focuses on social coding platforms (e.g., GitHub) which are a type of OSS platform. Besides version control features (e.g., fork, pull, and push features), social coding platforms have additional features for social interactions (e.g., starring and following). Collaboration occurs through the "fork and pull" model. A developer uses the "fork" feature to create a personal copy of an existing OSS project. Next, the "pull" feature allows all changes made on the personal copy to be merged with the source project. Traditional OSS platforms (e.g., SourceForge) do not support social coding. GitHub, however, records social interactions (e.g., following). As shown in Fig. 1, GitHub displays the number of "followers", "following", and "stars". GitHub API provides historical data with timestamps about these social interactions. In this study, social coding features in general, and starring and P2P following features in particular, are the boundary conditions for the findings. We expect the findings will apply to traditional OSS platforms if they implement social coding features including starring and P2P following features.

Data Collection

Our main data sources are ghtorrent.org and gharchive.org which provide archival GitHub data. In the process of determining the time range for our dataset, we refer to a recent study (Moqri et al. 2018) on GitHub which also included social network variables. As noted by their study, "the level of social interactions on the platform sharply declined by around 75% after December 2013" (p. 1199)[2]. Ideally, a current dataset

[2] The sharp decline in social interactions does not mean that individuals on GitHub reduced their social interactions. This occurs because the GitHub API stopped providing Follow events. We explain this in detail in Appendix A. Because only the API is affected,

would suit our study better. However, based on online documentation, GitHub API stopped providing data on the event "follow" at one point. We cross-checked the archive database provided by Google API and found that the follow events stopped in December 2013. In our study, we thus focus on projects initiated from January 2012 to December 2013. In Appendix A, we explain the relevance of our dataset. During our study period, we identify projects that receive no additional stars and whose initiators' social networks do not change (i.e., no additional followers and no additional followees). We remove these projects as they are irrelevant to our study. We randomly sample 5,500 unique project initiators. We ensure that the random sample is representative of the overall sample in terms of the number of stars, pull requests, and commits per project created. The total number of observations is 128,362[3].

Data Measurement

Dependent Variable

The dependent variable is an OSS initiator's project popularity. Following prior studies (Medappa and Srivastava 2019), we measure an OSS initiator's project popularity by counting the number of monthly new stars across all OSS projects created by the initiator. In our sample, the monthly average number of new stars per initiator is .33.

Independent Variables

We compute the characteristics of the social network of OSS initiators. On the one hand, we identify the follower network of each initiator by reporting the list of OSS community members who follow the initiator. A follower is an individual who can see the projects created or starred by the initiator. On the other hand, we create the followee network of each initiator by reporting the list of OSS community members that the initiator follows. The initiator receives information about the projects created or starred by the followees. The follower social network represents the direct followers of initiators while the followee social network characterizes the direct followees of initiators. Our two main independent variables are the numbers of an initiator's new followers and new followees. The *NewFollowers* variable represents the number of new direct

we expect that our findings still hold in 2021. The dataset is still relevant in 2021 because we do not observe any changes on the platform that would fundamentally influence starring and following activities (see Appendix A).

[3] We remove outliers based on the Cook's distance following the general rule of Cook's distance greater than 4/n (n = total number of observations). About 3.7% of the observations fall into the category of outliers.

followers while the *NewFollowees* variable is the number of new direct followees in a month. In our sample, the monthly average numbers of new followers and followees per initiator are .27 and 1.52 respectively, suggesting that on average initiators follow more OSS community members when compared to the number of OSS community members who follow them.

Moderating Variables

We compute the characteristics of the new followers' and followees' social networks. The monthly average numbers of followers among an initiator's new followers and followees represent the new follower and followee importance, respectively (see Equations 1 and 2).

$$NewFollowerImportance_{it} = \sum_{j=1}^{n_{it}} \frac{II_{ijt}}{n_{it}} \quad (1)$$

where the subscripts i, j, and t represent the initiator, a new follower, and the month, respectively. n_{it} is the number of new followers of the initiator i during the month t. II_{ijt} is the number of followers of the new follower j of the initiator i during the month t.

$$NewFolloweeImportance_{it} = \sum_{j=1}^{n_{it}} \frac{OI_{ijt}}{n_{it}} \quad (2)$$

where the subscripts i, j, and t represent the initiator, a new followee, and the month, respectively. n_{it} is the number of new followees of the initiator i during the month t. OI_{ijt} is the number of followers of the new followee j of the initiator i during the month t.

In our sample, an initiator's new followers and followees have on average 4.41 and 364.22 followers each, respectively. Hence, on average initiators follow more important OSS community members when compared to the importance of OSS community members who follow them.

The monthly average numbers of followees among an initiator's new followers and followees represent the new follower and followee reach, respectively (see Equations 3 and 4).

$$NewFollowerReach_{it} = \sum_{j=1}^{n_{it}} \frac{IO_{ijt}}{n_{it}} \quad (3)$$

where the subscripts i, j, and t represent the initiator, a new follower, and the month, respectively. n_{it} is the number of new followers of the initiator i during the month t. IO_{ijt} is the number of followees of the new follower j of the initiator i during the month t.

$$NewFolloweeReach_{it} = \sum_{j=1}^{n_{it}} \frac{OO_{ijt}}{n_{it}} \qquad (4)$$

where the subscripts i, j, and t represent the initiator, a new followee, and the month, respectively. n_{it} is the number of new followees of the initiator i during the month t. OO_{ijt} is the number of followees of the new followee j of the initiator i during the month t.

In our sample, an initiator's new followers and followees have on average 13.72 and 11.12 followees each, respectively. Hence, on average initiators attract the interest of OSS community members with higher reach when compared to the reach of OSS community members they follow.

Control Variables

We use four control variables. We measure the tenure of an initiator by counting the number of months since the initiator joined GitHub. Additionally, we control for the number of total projects created by an initiator. Moreover, we distinguish initiators associated with an organization (e.g., firm or OSS foundation) from those who are independent. Lastly, we include the number of bidirectional links of an initiator. A bidirectional link occurs when an initiator follows the follower. We describe the measurement of the variables in Table 3.

Table 3. Variable Measurement

Variable Name	**Variable Measurement**
OSS Project Popularity	Number of new stars across an initiator's OSS[4] in a month
New Followers	Number of new direct followers of an initiator in a month
New Leaders	Number of new direct followees of an initiator in a month
New Follower Reach	Average number of followees among an initiator's monthly new direct followers
New Follower Importance	Average number of followers among an initiator's monthly new direct followers
New Followee Reach	Average number of followees among an initiator's monthly new direct followees
New Followee Importance	Average number of followers among an initiator's monthly new direct followees
New Bidirectional Links	Number of new bidirectional links of an initiator in a month

[4] An OSS is the same as repository and project.

Tenure	Number of months since an initiator has registered on GitHub
Projects	Total number of projects an initiator has created by the end of a specific month
Company	Indicator, 1 if an initiator is enlisted in an organization on GitHub

Econometric Models

We assess the research model using the following econometric models (Equation 5 for the direct effects and Equation 6 for the moderating effects):

$$OSS\_Popularity_{i,t} = \beta_0 + \beta_1 NewFollowers_{i,t-1} + \beta_2 NewFollowees_{i,t-1} + \beta_3 NewBidirectionalLinks_{i,t-1} + \beta_4 Tenure_{i,t-1} + \beta_5 Projects_{i,t-1} + \beta_6 Company_{i,t-1} + v_i + \theta_t + \epsilon_{i,t} \quad (5)$$

$$OSS\_Popularity_{i,t} = \beta_0 + \beta_1 NewFollowers_{i,t-1} + \beta_2 NewFollowerReach_{i,t-1} + \beta_3 NewFollowerImportance_{i,t-1} + \beta_4 NewFollowers_{i,t-1} \times NewFollowerReach_{i,t-1} + \beta_5 NewFollowers_{i,t-1} \times NewFollowerImportance_{i,t-1} + \beta_6 NewFollowees_{i,t-1} + \beta_7 NewFolloweeReach_{i,t-1} + \beta_8 NewFolloweeImportance_{i,t-1} + \beta_9 NewFollowees_{i,t-1} \times NewFolloweeReach_{i,t-1} + \beta_{10} NewFollowees_{i,t-1} \times NewFolloweeImportance_{i,t-1} + \beta_{11} NewBidirectionalLinks_{i,t-1} + \beta_{12} Tenure_{i,t-1} + \beta_{13} Projects_{i,t-1} + \beta_{14} Company_{i,t-1} + v_i + \theta_t + \epsilon_{i,t} \quad (6)$$

where the subscripts i and t represent the initiator and the month, respectively. v_i are initiator fixed effects, and θ_t are month fixed effects. Equations 5 and 6 account for static time-invariant initiator heterogeneity and for seasonality and time trends.

Results

We report the descriptive statistics and the correlation matrix in Table 4. The dependent variable of interest is a count variable that is skewed and nonnegative. We therefore estimate the model using a panel negative binomial regression to account for the structure of the dependent variable. We estimate the coefficients of the econometric models using panel random effect and fixed effect estimators. In Table 5, we present the results of the fixed and random effects regression models. Models (1) and (3) estimate the direct effects of the number of an initiator's new followers and the number of an initiator's new followees on the initiator's project popularity. Models (2) and (4) assess the moderating effects of a new follower's reach and importance, and a new followee's reach and importance. According to the Hausman test, the fixed-effects model provides consistent estimates. Hence, we interpret the results of the fixed effects regression model. We use Model (3) to test H1 and H2, and Model (4) to test H1A, H1B, H2A, and H2B.

Table 4. Data Descriptive and Correlations

Variable	N	Mean	S.D.	Min	Max	1	2
1. OSS Project Popularity	128,362	.33	1	0	30	1	
2. NewFollowers	128,362	.27	1.17	0	83	-0.02*	1
3. NewFollowerReach	128,362	19.51	205.18	0	12,672	-0.01*	0.18*
4. NewFollowerImportance	128,362	3.49	67.85	0	9,296	-0.006*	0.12*
5. NewFollowees	128,362	1.52	5.7	0	433	0.07*	0.11*
6. NewFolloweeReach	128,362	10.18	46.74	0	3,991	0.002	0.09*
7. NewFolloweeImportance	128,362	302.2	2,353.11	0	80,612	-0.01*	0.03*
8. Tenure	128,362	20.4	13	1	51	0.03*	0.04*
9. Projects	128,362	.73	.56	0	7	0.24*	-0.03*
10. NewBidirectionalLinks	128,362	.04	.23	0	9	-0.02*	0.5*

* p<0.05.

Variable	3	4	5	6	7	8	9
3. NewFollowerReach	1						
4. NewFollowerImportance	0.09*	1					
5. NewFollowees	0.04*	0.02*	1				
6. NewFolloweeReach	0.05*	0.04*	0.09*	1			
7. NewFolloweeImportance	0.02*	0.02*	0.02*	0.39*	1		
8. Tenure	0.01*	0.006*	0.14*	0.1*	0.05*	1	
9. Projects	-0.05*	-0.02*	0.02*	-0.003	0.03*	0.1*	1
10. NewBidirectionalLinks	0.05*	0.1*	0.11*	0.07*	0.01*	-0.002	-0.05*

* p<0.05.

Table 5. Negative Binomial Regression Models

	Random Effects		Fixed Effects	
	Model 1	**Model 2**	**Model 3**	**Model 4**
Lag1.NewFollowers	.0152*	.0162*	.0133*	.0153*
	(.0061)	(.0069)	(.006)	(.0068)
Lag1.NewFollowerReach		.0031		.0012
		(.0083)		(.0083)
Lag1.NewFollowerImportance		-.0073		-.0056
		(.0112)		(.0109)
Lag1.NewFollowers*NewFollowerReach		-.0091		-.0075
		(.0111)		(.011)
Lag1.NewFollowers*NewFollowerImportance		.0015		-6.4e-04
		(.0077)		(.0079)
Lag1.NewFollowees	.204***	.201***	.225***	.216***
	(.0151)	(.0159)	(.0169)	(.0176)
Lag1.NewFolloweeReach		.01*		.0096*
		(.0043)		(.0043)
Lag1.NewFolloweeImportance		.0083		.0089
		(.007)		(.0071)
Lag1.NewFollowees*NewFolloweeReach		.0547***		.0563***
		(.0106)		(.0107)
Lag1.NewFollowees*NewFolloweeImportance		.136*		.125*
		(.0621)		(.0627)
Lag1.Tenure	.0034**	.0033**		
	(.0011)	(.0011)		
Lag1.Projects	1.34***	1.35***	1.37***	1.38***
	(.0196)	(.0197)	(.0217)	(.0217)

Lag1.NewBidirectionalLinks	.0104	.005	.0164+	.0112
	(.0096)	(.0098)	(.0099)	(.0101)
Lag1.Company	-.0266	-.0252		
	(.0303)	(.0304)		
Constant	-1.68***	-1.69***	-1.65***	-1.66***
	(.0605)	(.0606)	(.0456)	(.0456)
Chi2	6894.59***	6942.09***	6157.17***	6201.35***

+ p <.1 * p<0.05, ** p<0.01, *** p<0.001.
Standardized coefficients included except for projects and tenure; the dependent variable is not standardized.
Month dummies included.

In hypothesis 1, we argue that the number of an initiator's new followers has a positive effect on the initiator's project popularity. The results of Model 3 (see Table 5) indicate that the number of an initiator's new followers has a statistically significant effect on the initiator's project popularity. Hence, hypothesis 1 is supported. A standard deviation (i.e., 1.17 units) increase in an initiator's number of new followers is associated with an increase in new stars by .013, corresponding to a 4.8% increase compared to the monthly average number of new stars. Moreover, in hypothesis 2, we posit that the number of an initiator's new followees has a positive effect on the initiator's project popularity. The results of Model 3 (see Table 5) show that the number of an initiator's new followees has a statistically significant effect on OSS project popularity. Hence, hypothesis 2 is supported. A standard deviation (i.e., 5.7 units) increase in an initiator's number of new followees is associated with an increase in new stars by .225, corresponding to a 14.8% increase compared to the monthly average number of new stars. Our results indicate that the effect of the number of an initiator's new followees on the initiator's project popularity is stronger than that of the initiator's new followers. We can conclude that the number of new followees is more important than the number of followers at increasing an initiator's project popularity.

In hypotheses 1A and 1B, we propose that new follower reach and new follower importance negatively influence the effect of the number of an initiator's new followers on the initiator's project popularity. We find no support for both hypotheses. The interactions between the number of an initiator's new followers and new follower reach as well as importance are not statistically significant. Our results suggest that an initiator's OSS project popularity does not depend on the ability of the initiator to attract followers of different importance and reach on open source platforms. Furthermore, in hypotheses 2A and 2B, we posit that new followee reach and new followee importance positively influence the effect of the number of an initiator's new followees on the initiator's project popularity. We find support for both hypotheses. The interactions between the number of an initiator's new followees and new followee reach as well as importance respectively are positive and statistically significant (see Model 4, Table 5). At a lower level of new followee reach (i.e., new followee reach equals 0 corresponding to 69% of the sample), a standard deviation (i.e., 5.7 units) increase in an initiator's number of new followees is associated with an increase in new stars by 0.206. When new followee reach is high (i.e., new followee reach equals 44.33 corresponding to 95 percentile), a standard deviation increase in an initiator's number of new followees is associated with an increase in new stars by 0.24. Moreover, at a lower level of new followee importance (i.e., new followee importance equals 0 corresponding to 65% of the sample), a standard deviation (i.e., 5.7 units) increase in an initiator's number of new followees is associated with an increase in new stars by 0.2. When new follower importance is high (i.e., new followee importance equals 449 corresponding to 95 percentile), a standard deviation increase in an initiator's number of new followees is associated with an increase in new stars by 0.22. Our results suggest that an OSS initiator's project popularity depends on the ability of the initiator to follow individuals who have more importance and reach on open source platforms. Figure 5 and Figure 6 illustrate the moderation effects[5].

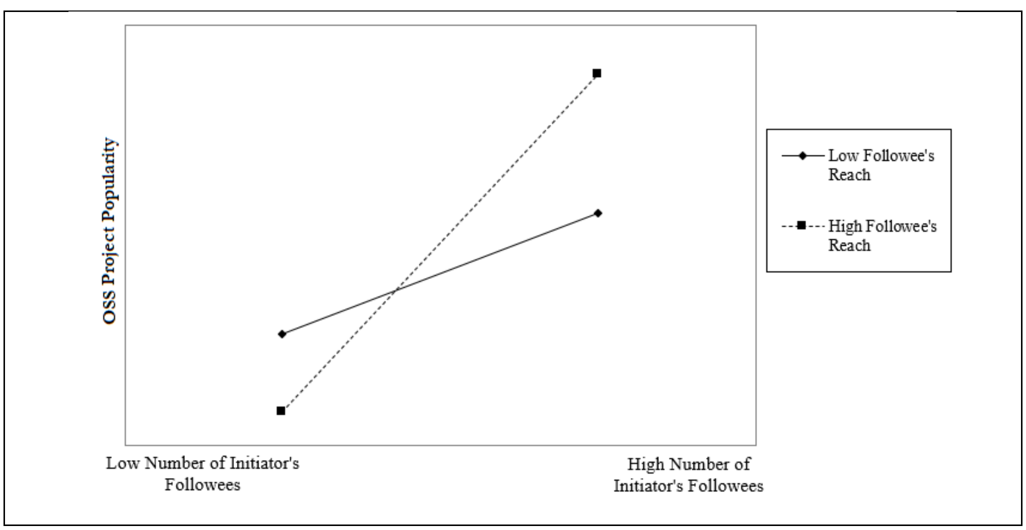


Figure 5. Effect of the Number of an Initiator's New Followees across Different Levels of New Followee Reach

[5] We create the moderation plots using resources developed by Jeremy Dawson (http://www.jeremydawson.co.uk/slopes.htm). Since our model has control variables, only the pattern of the moderation is interpretable. The values of the OSS project popularity (y-axis) are inaccurate. Hence, we are not reporting the values of the y-axis.

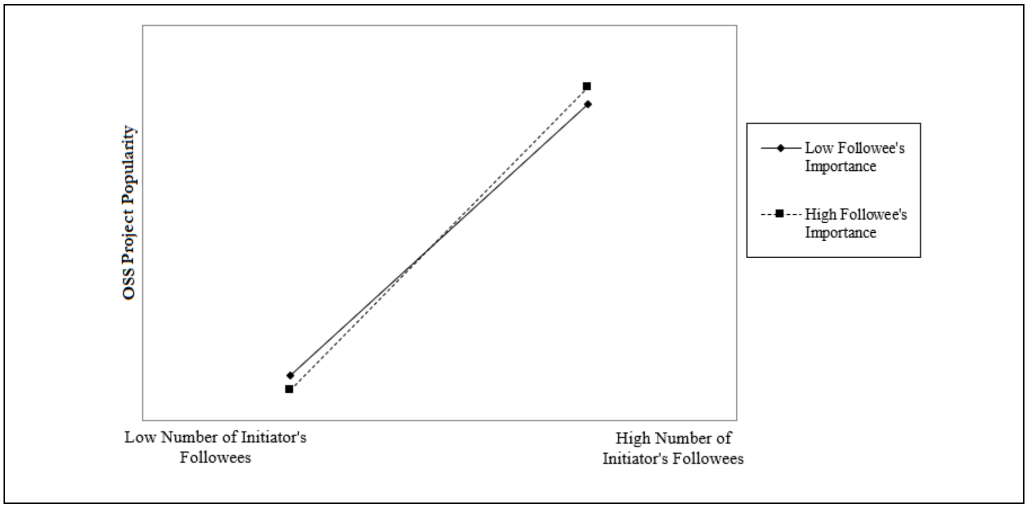


Figure 6. Effect of the Number of an Initiator's New Followees across Different Levels of New Followee Importance

Endogeneity

Considering the length of our panel, we calculate estimates from the Arellano-Bond estimator, a generalized method of moments (GMM)-based estimator. The Arellano-Bond System GMM is a dynamic panel estimator that accounts for endogenous regressors by using lagged differences as instruments (Moqri et al. 2018). The results of the Arellano-Bond estimator, shown in Table 6, are consistent with the main findings, indicating that endogeneity is not a major concern in our analyses.

Table 6. Arellano-Bond Models with Robust Standard Errors

	Arellano Bond	
	Model 1	**Model 2**
Lag1.NewStars	.171***	.213***
	(.0084)	(.0084)
NewFollowers	.0138***	.0131**
	(.0038)	(.0043)
NewFollowerReach		5.0e-06
		(.0017)
NewFollowerImportance		8.6e-04
		(.0022)
NewFollowers*NewFollowerReach		6.3e-04
		(.0018)
NewFollowers*NewFollowerImportance		.0014
		(.0017)
NewFollowees	.101***	.0714***
	(.0126)	(.0091)
NewFolloweeReach		.0118***
		(.0019)
NewFolloweeImportance		-.002
		(.0015)

NewFollowees*NewFolloweeReach		.0278*
		(.0118)
NewFollowees*NewFolloweeImportance		.0452*
		(.0177)
Projects	.222***	.15***
	(.0081)	(.0046)
NewBidirectionalLinks	-.0027	-.0059**
	(.0018)	(.0019)
Chi2	1944.34***	3016.65***

+ p <.1 * p<0.05, ** p<0.01, *** p<0.001.
Standardized coefficients included except for projects and tenure; the dependent variable is not standardized.
Month dummies include.

Moreover, we use a panel vector autoregressive (PVAR) model to account for the endogeneity of the number of new stars, new followers, and new followees. PVAR is a robust dynamic panel estimator that accounts for unobserved individual heterogeneity and models each endogenous variable as a function of its lagged values and other endogenous variables' lagged values to calculate consistent estimates. The PVAR model controls for time effects, individual fixed effects, and time-dependent covariates. To ensure that the estimates are robust, we use the Helmert transformation, a forward differencing method that controls individual fixed effects. Following prior literature, we estimate several models using the first three lags of the number of new stars, new followers, and new followees as "GMM-style" instruments (Moqri et al. 2018). The results, shown in Table 7, indicate that the first-order panel VAR model is more appropriate for the analyses. Notably, this model has the lowest MBIC and MQIC. Additionally, we ensure that the first-order panel VAR model satisfies the stability condition. Table 8 and Figure 7 indicate that the model is stable as all the eigenvalues are less than one. The results of the first-order panel VAR model are presented in Table 9. The first lag of the number of new followees has a positive effect on the number of new stars. However, the first lag of the number of new stars does not influence the number of new followees.

Table 7. Panel VAR Model Selection

lag	CD	J	J p-value	MBIC	MAIC	MQIC
1	.9578143	40.53163	.0456547	-265.4066	-13.46837	-90.55587
2	.9248014	13.89778	.7357298	-190.061	-22.10222	-73.49388
3	.7152739	6.004754	.7394424	-95.97464	-11.99525	-37.69108

Notes. CD: Coefficient of Determination; J: Hansen's J; MBIC: Bayesian Information Criterion; MAIC: Akaike Information Criterion; MQIC: Hannan-Quinn Information Criterion.

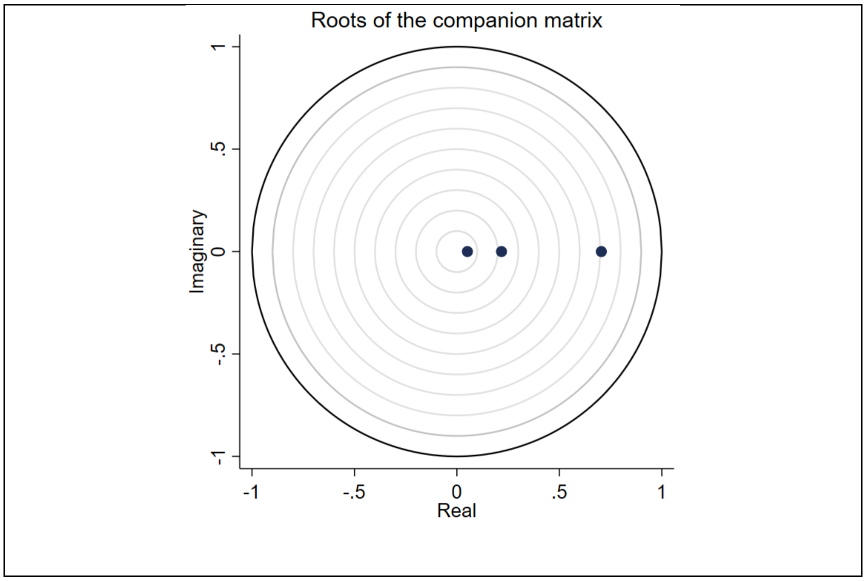


Figure 7. Eigenvalues Inside the Unit Circle

Table 8. Eigenvalue Stability Condition

Eigenvalue		
Real	**Imaginary**	**Modulus**
.7049471	0	.7049471
.2177506	0	.2177506
.0513103	0	.0513103

Table 9. Panel VAR Estimation for New Stars with Robust Standard Errors

Independent Variable	**Dependent Variable**		
	NewStars	**NewFollowers**	**NewFollowees**
Lag1.NewStars	.196***	-2.9e-04+	-2.5e-04
	(.0559)	(1.5e-04)	(4.0e-04)
Lag1.NewFollowers	-.0542	.0876+	.005*
	(.0391)	(.0527)	(.0023)
Lag1.NewFollowees	.888***	-2.8e-04	.776***
	(.246)	(.0071)	(.0415)

+ p <.1 * p<0.05, ** p<0.01, *** p<0.001.
The numbers of new followers and followees are standardized; the number of stars is not standardized.

Furthermore, we determine the Granger causality to alleviate concerns about reverse causality (Moqri et al. 2018). Conceptually, we do think there is reverse causality in our analyses. The number of stars across an initiator's OSS could benefit any project's members and not just the initiator. Moqri et al. (2018) highlight that heavy contributors attract more attention from users. Hence, unless being a heavy contributor, an initiator might not benefit from an increasing number of stars. Moreover, starers might not have an interest in an initiator's other activities. A starrer's interest might be limited to a specific initiator's OSS and does not include other OSS. Thus, receiving more stars on a particular OSS does not always mean that an initiator will receive new followers. Additionally, the number of followees that an initiator follows indicates the initiator's interest in the followees' activities (i.e., starring and project creation). It is not clear how the number of stars might influence an initiator's following behavior. Empirically, the Granger causality results (see Table 10) suggest that the numbers of new followees Grander-cause the number of new stars. However, the opposite is not true.

Table 10. Panel VAR-Granger Causality Wald Test

Equation\	Excluded	chi2	df	Prob > chi2
NewStars	NewFollowers	1.926	1	0.165
	NewFollowees	13.058	1	0.000
NewFollowers	NewStars	3.684	1	0.055
	NewFollowees	0.002	1	0.969
NewFollowees	NewStars	0.385	1	0.535
	NewFollowers	4.758	1	0.029

Robustness Checks

Having used a panel negative binomial regression in the main analyses, we conduct another estimation using a panel Poisson model. The obtained standard errors are robust to arbitrary patterns of serial correlation, therefore making them well-suited to address issues of autocorrelation in time-series models. The results are mostly consistent with our main findings, except for the effect of the number of new followers that is not significant (see Table 11). We also show that our main findings are robust to omitted variables. We incorporate additional explanatory variables including the total number of contributions by project members, an initiator's total numbers of affiliated repositories, an initiator's number of contributions, and the total number of pull requests by project members. The results (see Table 12) corroborate the main findings.

Table 11. Poisson Models with Robust Standard Errors

	Random Effects		Fixed Effects	
	Model 1	Model 2	Model 3	Model 4
Lag1.NewFollowers	.0115	.0142	.0099	.0136
	(.009)	(.0102)	(.0099)	(.0113)
Lag1.NewFollowerReach		-.0027		-.0038
		(.007)		(.007)
Lag1.NewFollowerImportance		-.0127		-.0102
		(.0113)		(.0101)
Lag1.NewFollowers*NewFollowerReach		-.0083		-.0072
		(.0115)		(.0115)
Lag1.NewFollowers*NewFollowerImportance		-.0023		-.0045
		(.0076)		(.0081)
Lag1.NewFollowees	.222***	.218***	.234***	.226***

	(.0491)	(.0419)	(.0562)	(.0485)
Lag1.NewFolloweeReach		.0111*		.0107*
		(.005)		(.0049)
Lag1.NewFolloweeImportance		.0097		.0106
		(.008)		(.0081)
Lag1.NewFollowees*NewFolloweeReach		.0441*		.0442*
		(.0186)		(.0185)
Lag1.NewFollowees*NewFolloweeImportance		.179*		.173*
		(.0792)		(.0801)
Lag1.Tenure	.002	.0019	.065***	.0839***
	(.0015)	(.0015)	(.0101)	(.0169)
Lag1.Projects	1.29***	1.3***	1.3***	1.31***
	(.0618)	(.0604)	(.0687)	(.0671)
Lag1.NewBidirectionalLinks	.0077	.0039	.0142	.0108
	(.0124)	(.0125)	(.0127)	(.0129)
Lag1.Company	-.232***	-.231***		
	(.0427)	(.0426)		
Chi2	7007.3***	7112.77***	2989.7***	2991.37***

+ p <.1 * p<0.05, ** p<0.01, *** p<0.001.
Standardized coefficients included except for projects and tenure; the dependent variable is not standardized.
Month dummies included.

Table 12. Negative Binomial Regression Models with Additional Variables

	Random Effects		**Fixed Effects**	
	Model 1	**Model 2**	**Model 3**	**Model 4**
Lag1.NewFollowers	.0152*	.016*	.0134*	.0153*
	(.0062)	(.0069)	(.0061)	(.0069)
Lag1.NewFollowerReach		.0029		.001
		(.0083)		(.0082)
Lag1.NewFollowerImportance		-.0078		-.0059
		(.0113)		(.0109)
Lag1.NewFollowers*NewFollowerReach		-.0091		-.0074
		(.0111)		(.011)
Lag1. NewFollowers*NewFollowerImportance		.002		-3.5e-04
		(.0077)		(.0078)
Lag1.NewFollowees	.2***	.198***	.222***	.213***
	(.0151)	(.0159)	(.0168)	(.0176)
Lag1.NewFolloweeReach		.0099*		.0095*
		(.0043)		(.0043)
Lag1.NewFolloweeImportance		.0086		.0093
		(.007)		(.0071)
Lag1.NewFollowees*NewFolloweeReach		.0541***		.0557***
		(.0106)		(.0107)
Lag1.NewFollowees*NewFolloweeImportance		.137*		.129*
		(.0621)		(.0628)
Lag1.Tenure	.0034**	.0034**		
	(.0011)	(.0011)		
Lag1.Projects	1.31***	1.32***	1.35***	1.36***
	(.0197)	(.0198)	(.0218)	(.0219)
Lag1.BidirectionalLinks	.0093	.004	.0157	.0105
	(.0096)	(.0098)	(.0099)	(.0101)

Lag1.Company	-.0257	-.024		
	(.0304)	(.0304)		
Lag1.Contributions	4.9e-05	4.6e-05	-4.7e-05	-5.0e-05
	(1.4e-04)	(1.4e-04)	(1.5e-04)	(1.5e-04)
Lag1.Affiliation_Repositions	1.4e-04	1.2e-04	1.2e-04	1.1e-04
	(1.3e-04)	(1.3e-04)	(1.5e-04)	(1.5e-04)
Lag1.Initiator_Contributions	.0045***	.0045***	.0042***	.0042***
	(4.6e-04)	(4.6e-04)	(4.9e-04)	(4.9e-04)
Lag1.Pull_Requests	.0071***	.007***	.0044**	.0044**
	(.0013)	(.0013)	(.0014)	(.0014)
Chi2	7083.09***	7130.07***	6264.33***	6308.06***

+ p <.1 * p<0.05, ** p<0.01, *** p<0.001.
Standardized coefficients included except for contributions, affiliated repositories, initiator's contributions, pull requests, projects, and tenure; the dependent variable is not standardized.
Month dummies included.

Following prior studies (e.g., Moqri et al. 2018), we use one-month lagged values in the models to estimate the effects of the predictors on the dependent variable. While one may argue that the predictors may have short and long-term effects on the dependent variable, our study follows prior research (e.g., Moqri et al. 2018) by focusing on the short-term impact of the predictors on the dependent variable. Lagged values of the predictors help reduce the issue of reverse causality. We conduct more analyses with no lagged values, and two months lagged values. The results presented in Tables 13 and 14 indicate that our results are robust when we use no lagged values, and two months lagged values, respectively.

Table 13. Negative Binomial Regression Models with No Lagged Values

	Random Effects		**Fixed Effects**	
	Model 1	**Model 2**	**Model 3**	**Model 4**
NewFollowers	.0186***	.0257	.0177***	.0379+
	(.0052)	(.0195)	(.0051)	(.02)
NewFollowerReach		-.0297*		-.0305*
		(.0138)		(.0136)
NewFollowerImportance		-.0249		-.0263
		(.0277)		(.0265)
NewFollowers*NewFollowerReach		-.0026		-.0053
		(.0149)		(.0151)
NewFollowers*NewFollowerImportance		.0154		.0249
		(.0226)		(.0246)
NewFollowees	.272***	.264***	.306***	.294***
	(.0162)	(.0182)	(.0177)	(.02)
NewFolloweeReach		.0198*		.0204*
		(.0098)		(.01)
NewFolloweeImportance		-.0173+		-.0151+
		(.009)		(.0091)
NewFollowees*NewFolloweeReach		.0714***		.0713**
		(.0215)		(.022)

NewFollowees*NewFolloweeImportance		.195**		.2**
		(.0718)		(.0724)
Tenure	.0011	.0013		
	(.0012)	(.0011)		
Projects	1.92***	1.95***	2.09***	2.12***
	(.0227)	(.0223)	(.0252)	(.0248)
NewBidirectionalLinks	-.0179	-.0304*	-.0123	-.0275*
	(.0115)	(.0127)	(.0117)	(.0128)
Company	.0055	.008		
	(.0325)	(.0323)		
Constant	-1.96***	-2.1***	-2.22***	-2.16***
	(.0629)	(.0638)	(.0796)	(.0479)
Chi2	9773.21***	9823.49***	9368.28***	9421.35***

+ p <.1 * p<0.05, ** p<0.01, *** p<0.001.
Standardized coefficients included except for projects and tenure; the dependent variable is not standardized.
Month dummies included.

Table 14. Negative Binomial Regression Models with Two Months Lagged Values

	Random Effects		Fixed Effects	
	Model 1	**Model 2**	**Model 3**	**Model 4**
Lag2.NewFollowers	.0217***	.0091	.0218***	.0091
	(.005)	(.0098)	(.005)	(.0095)
Lag2.NewFollowerReach		.006		.0045
		(.0047)		(.0048)
Lag2.NewFollowerImportance		.0141*		.0127+
		(.0069)		(.007)
Lag2.NewFollowers*NewFollowerReach		.0064		.0073
		(.0103)		(.0099)
Lag2.NewFollowers*NewFollowerImportance		.0019		.0016
		(.0122)		(.0123)
Lag2.NewFollowees	.213***	.116***	.229***	.125***
	(.0153)	(.0228)	(.017)	(.0235)
Lag2.NewFolloweeReach		-.0055		-.0067
		(.0091)		(.009)
Lag2.NewFolloweeImportance		-.0131		-.0126
		(.0137)		(.0134)
Lag2.NewFollowees*NewFolloweeReach		.0658***		.0697***
		(.0145)		(.0148)
Lag2.NewFollowees*NewFolloweeImportance		.195**		.176**
		(.0613)		(.062)
Lag2.Tenure	.0033**	.0032**		
	(.0011)	(.0011)		
Lag2.Projects	1.12***	1.12***	1.1***	1.11***
	(.0194)	(.0194)	(.0216)	(.0216)
Lag2.NewBidirectionalLinks	.0123	.0086	.0202*	.0157
	(.0087)	(.0096)	(.009)	(.0099)
Lag2.Company	-.0394	-.0393		
	(.0301)	(.0301)		
Constant	-1.4***	-1.41***	-1.47***	-1.5***
	(.0578)	(.058)	(.0703)	(.0706)
Chi2	5442.49***	5479.7***	4675.04***	4710.89***

+ p <.1 * p<0.05, ** p<0.01, *** p<0.001.
Standardized coefficients included except for projects and tenure; the dependent variable is not standardized.
Month dummies included.

Furthermore, we collect additional data to complement the data used in the main analyses. The new dataset covers the year 2016-2017 while the original dataset covers 2012-2013. One may argue that the 2012-2013 dataset is likely not relevant to today's open source community practices. Although we do not see any major reasons to believe that the 2012-2013 dataset is no longer relevant (i.e., the features to create followership networks, and star projects have not changed since the start date of data collection), we still conduct additional analyses using the 2016-2017 (January 2016 – January 2017) dataset to alleviate any concerns related to the relevance of the dataset. Unfortunately, the 2016-2017 dataset is incomplete because GitHub API stopped providing Follow events after December 2013. Thus, the exact date of the creation of the Follow events may not be accurate. GHTorrent states that "the *created_at* field is only filled in accurately for followerships for which GHTorrent has recorded a corresponding event. Otherwise, it is filled in with the latest date that the corresponding user or follower has been created." (ghtorrent.org/relational.html). Hence, the additional dataset (January 2016 – January 2017) presents this limitation. And although the results (see Table 15) corroborate our main findings and seem to reduce concerns about the validity of our main findings, we believe that the results should be interpreted in light of the current limitations of the dataset. Finally, we show that the results do not change after including additional control variables such as the average project age, the number of programming languages, and the total project member per initiator (see Table 16).

Table 15. Negative Binomial Regression Models with an Additional Dataset

	Random Effects		Fixed Effects	
	Model 1	**Model 2**	**Model 3**	**Model 4**
Lag1.NewFollowers	.021***	.0829	.0208***	.0723
	(.0051)	(.0638)	(.005)	(.0639)
Lag1.NewFollowerReach		-.0204		-.0213
		(.097)		(.0984)
Lag1.NewFollowerImportance		.0318		.0337
		(.103)		(.103)
Lag1.NewFollowers*NewFollowerReach		-.271		-.272
		(.237)		(.234)
Lag1.NewFollowers*NewFollowerImportance		.272		.266
		(.237)		(.232)
Lag1.NewFollowees	.368*	.208	.358*	.689
	(.175)	(.1688)	(.175)	(.487)
Lag1.NewFolloweeReach		-.239**		-.205+

		(.0807)		(.121)
Lag1.NewFolloweeImportance		-.0126		-.0188
		(.0109)		(.0133)
Lag1.NewFollowees*NewFolloweeReach		.202***		.179*
		(.0608)		(.0863)
Lag1.NewFollowees*NewFolloweeImportance		.982***		.938***
		(.236)		(.24)
Lag1.Tenure	.0096***	.0035		
	(7.7e-04)	(.0028)		
Lag1.Projects	.35***	1.54***	.354***	1.59**
	(.018)	(.461)	(.0177)	(.54)
Lag1.NewBidirectionalLinks	-.0575+	-.0908*	-.0595+	-.0604
	(.0315)	(.041)	(.0318)	(.0668)
Lag1.Company	-.0736+	.347*		
	(.0389)	(.152)		
Constant	-.1	-.49***	.383***	.235*
	(.0654)	(.0655)	(.0412)	(.11)
Chi2	1026.47***	1054.88***	792.34***	825.68***

N = 66,696
+ p <.1 * p<0.05, ** p<0.01, *** p<0.001.
Standardized coefficients included except for projects and tenure; the dependent variable is not standardized.
Month dummies included.

Table 16. Negative Binomial Regression Models with Additional Control Variables

	Random Effects		Fixed Effects	
	Model 1	**Model 2**	**Model 3**	**Model 4**
Lag1.NewFollowers	.0222***	.0539	.0216***	.0536
	(.005)	(.038)	(.005)	(.038)
Lag1.NewFollowerReach		.0171		.0162
		(.0728)		(.0729)
Lag1.NewFollowerImportance		-.0169		-.0163
		(.0803)		(.0803)
Lag1.NewFollowers*NewFollowerReach		-.258		-.255
		(.189)		(.189)
Lag1.NewFollowers*NewFollowerImportance		.283		.281
		(.193)		(.193)
Lag1.NewFollowees	.335*	.194**	.338*	.194**
	(.17)	(.0698)	(.17)	(.0696)
Lag1.NewFolloweeReach		-.217**		-.217**
		(.0821)		(.0815)
Lag1.NewFolloweeImportance		-.0107		-.0103
		(.0099)		(.01)
Lag1.NewFollowees*NewFolloweeReach		.181**		.18**
		(.0613)		(.061)
Lag1.NewFollowees*NewFolloweeImportance		.923***		.918***
		(.231)		(.231)
Lag1.Tenure	.0036***	.0104*	-1.6e-04	.0066
	(8.5e-04)	(.0049)	(.0012)	(.0051)
Lag1.Projects	.174***	1.59**	.177***	1.6**
	(.0218)	(.506)	(.022)	(.507)
Lag1.NewBidirectionalLinks	-.0529+	-.0751*	-.0532+	-.0733*

	(.0315)	(.0359)	(.0316)	(.036)
Lag1.Company	-.0284	.138	-.037	.124
	(.0397)	(.176)	(.0398)	(.181)
Lag1.AvgLanguages	.294***	-.885	.282***	-.893
	(.0468)	(.542)	(.0471)	(.544)
Lag1. ProjectAge	.307***	-.114	.295***	-.12
	(.019)	(.165)	(.0192)	(.166)
Lag1. CumMembers	.12***	-.303	.111***	-.311
	(.0284)	(.217)	(.0285)	(.217)
Constant	-.852***	-.376	.172	-.0617
	(.0872)	(.664)	(.259)	(.765)
Chi2	2260.81***	2278.56***	1199.87***	1200.76***

N = 66,696; + p <.1 * p<0.05, ** p<0.01, *** p<0.001.
Standardized coefficients included except for total numbers of members across initiator's projects, languages, project's age, projects, and tenure; the dependent variable is not standardized.
Month dummies included.

Discussion

Theoretical Contributions

This study is among the first to examine the effect of P2P followership on OSS initiators' project popularity. Following social media platforms such as Facebook, Twitter, or Instagram, we reveal the significance of P2P followership on OSS development platforms such as GitHub. Previous research has been limited to affiliation and communication networks (Peng et al. 2013; Sutanto et al. 2014, 2021) which do not capture the idea of P2P followership. In those networks, individuals are connected because they are either members of the same project or collaborating in content creation. As we unravel the influence of project initiators' P2P followership networks on their project popularity, we extend prior research on OSS project popularity. Moreover, by shedding light on the role of both the quantity and connectivity of OSS initiators' P2P followership nodes, we extend prior studies that have not highlighted the connectivity-related heterogeneity of P2P followership nodes in OSS communities (Jiang et al. 2019; Moqri et al. 2018). Our findings highlight the necessity of integrating both the quantity and connectivity of P2P followership nodes to understand an initiator's project popularity on an OSS development platform.

Moreover, this study extends prior literature by attempting to connect social media literature with the OSS literature. By doing so, this work suggests a different conceptualization of OSS development that incorporates two main social coding platform capabilities: first, content creation and access; and second, the establishment and management of P2P followership. Recently, research on OSS development begins to consider the effect of P2P followership (e.g., Jiang et al. 2019; Moqri et al. 2018). The theoretical implication of this effort can be extended to the context of online production communities. Those two capabilities offer a better theoretical framework to understand not only OSS development success but also online production communities' effectiveness in general. This study intends to stimulate more research on the connection between social media and the online production community. We believe that the social media perspective can provide researchers with new insights for studying online production communities.

Practical Implications

A rich and diverse ecosystem evolves around OSS development including platform owners, project initiators, contributors, and end-users. All these actors are concerned with the takeoffs of OSS projects which are beneficial for all. In this ecosystem, platform owners create the platform capabilities through which project initiators create new projects, contributors participate in the development of such projects, and end-users adopt the finished products. Looking at takeoffs in terms of project popularity within the OSS community, it is important for us to understand the factors at play. Our findings have implications for both the platform owners and project initiators in the OSS community.

This study reveals that P2P followership influences an OSS initiator's project popularity. More specifically, P2P followership and network characteristics affect an initiator's project popularity. The networks of followers and followees have a positive impact on an initiator's project popularity within the OSS community. Such results are representative of the growing role of P2P followership on OSS development platforms. OSS studies have been limited to the creation and access of content without an emphasis on social networks. This study implies that platform owners should develop more P2P followership capabilities within the OSS community. By enabling more socialization among the OSS community, OSS development platforms such as GitHub are facilitating the popularization of OSS projects. Thus, platform owners who want to unlock the power of OSS should also consider offering social network features.

This study shows that project initiators with high influence benefits within the OSS community are those who attract many followers, and they are more likely to develop successful projects. The influence benefits embedded in the structure of the follower network enable project initiators to achieve high project popularity in the OSS community. By examining the network of project initiators' followers and the network characteristics of their followers, we gain ideas on the influence benefits of project initiators within the OSS community. Individuals with such influence benefits are more visible within the community, and therefore the effect of their participation is widespread. People with high influence benefits are those who have a large number of followers in the OSS community. Project initiators can be individuals affiliated with organizations, or independent individuals. As initiators are more likely leaders of OSS projects, it becomes extremely important for them to leverage their effort and time in OSS development by building more influence benefits. Surprisingly, we do not find support for a moderating effect of follower importance and reach on the relationship between follower network and an initiator's project popularity. We build our theoretical model based on the argument that followers are different and those that are tightly connected with an initiator are more likely to be influenced. However, our results suggest that the number of followers positively predicts an OSS initiator's project popularity, and the follower heterogeneity does not make a difference. Several plausible explanations could justify the lack of significance. For example, it could be argued that, although followers with low reach and importance may be tightly connected (i.e., give more time and attention) to the initiator, they are not easily influenced when compared to followers with high reach and importance. Notably, low reach and importance may mean less interest in OSS projects. Another reason is that high reach and importance have two opposing effects that cancel each other. On one hand, high reach and importance may give less time and attention to the initiator (i.e., loose connections) and thus dilute the influence effect of the follower network. On the other hand, high reach and importance may help draw more time and attention from the rest of the community (i.e., tight connections) and thus enhance the influence effect of the follower network. Hence, more research is still needed to better our understanding of the influence of different types of followers.

This study demonstrates that project initiators who have access to large and diverse sources of information within the OSS community are more likely to develop successful projects. The information benefits embedded in the structure of the followee network allow project initiators to achieve high popularity in the OSS community. Access to large and diverse sources of information is considered as an information benefit because it opens doors for opportunities that would not have been made possible otherwise. By examining the network of people project initiators are following and the importance of the people they follow, we gain ideas on the information benefits of project initiators within the OSS community. First, the action of following people within the OSS community provides project initiators the opportunity to access rich information in terms of OSS development activities. It develops the level of awareness or sensing within the community. Project initiators with such opportunities benefit from the information they have access to and can generate more attractive projects from a community perspective. Such information benefits may come from different sources among which the number of people initiators are following is the focal one that we are studying. This study, therefore, emphasizes the importance for project initiators to consider information benefits to leverage their effort and time on OSS development.

Our study further suggests that following individuals should not be random, and the network can be strategically managed by project initiators. Social networks can reflect certain strategic orientations. For example, a unified network orientation positively relates to innovation involvement (Obstfeld 2005). Our study points out a direction that may potentially help project initiators benefit from high quality following, i.e., by following people with high importance and high reach. On OSS development platforms, information is abundant, and following should be selective and wise. Otherwise, more time and effort may be wasted on information searching. People nowadays are burdened with information overload rather than information shortage. Project initiators targeting high popularity should strike the balance between staying sufficiently informed and reducing information overload. Therefore, mindful use of time on quality information becomes extremely important. Our study provides a possible venue for achieving that.

Limitations and Future Direction

While our study presents unique findings on the role of OSS project initiators' social networks in enhancing their project popularity, four main limitations may require further attention in future studies. First, our dataset is limited to 2012 – 2013 because GitHub API stops providing Follow events after December 2013. We do ensure that the 2012 – 2013 dataset is relevant by showing that the following: (a) it is unique; (b) it has been used in prior studies (Moqri et al. 2018); (c) newer datasets are incomplete; and (d) there are no changes on the platform that fundamentally influence starring and following activities (see Appendix A). However, future studies might explore ways to collect newer datasets. Second, our study is limited to the structural aspects of social networks as we investigate the effect of various aspects of a project initiator's social network structure on the initiator's project popularity. Social networks also encompass relational and cognitive dimensions (Nahapiet and Ghoshal 1998). Future studies can explore these other dimensions in social networks and their interactions with the structural dimension. Third, we provide several analyses, including Arellano-Bond dynamic panel estimations to address endogeneity issues in our findings. A better assessment of causality claims necessitates more controlled environments through experiments or quasi-experiments. Fourth, we focus on a limited number of metrics while studying initiators' follower and followee social network structures. We measure the reach and importance of an initiator's followers and followees by relying on the number of followees and followers. Similarly, we assess an OSS project initiator's social network structure by calculating the number of followees and followers of the initiator. Future research can explore additional social network metrics including closeness centrality, betweenness centrality, and eigenvector centrality. Besides the adoption of advanced measures of the social network structure, future studies might explore additional OSS outcomes such as code reuse, number of forks, and number of commits.

Our study creates opportunities for future research by highlighting the significance of understanding the role of a project initiator's social networks in increasing the initiator's project popularity. We propose that future research may explore contingent factors such as market uncertainty (i.e., dynamism, munificence, and competition). Moreover, our findings can be extended to non-initiators. Future studies may investigate how OSS project members' social network structure contributes to the overall popularity of the project.

Conclusion

P2P followership is prevalent among OSS development platforms. Understanding the impact of project initiators' P2P followership networks on their project popularity requires an exploration of both the quantity and connectivity of project initiators' P2P followership nodes. We empirically investigate the effect of OSS project initiators' P2P followership networks on their project popularity using a large panel dataset over 24 months from January 2012 to December 2013. Our research unravels the heterogeneity among an initiator's P2P followership nodes while highlighting the information and influence benefits that an initiator derives from the P2P followership nodes. Our findings provide nuanced insights into how an OSS project initiator's P2P followership nodes impact the initiator's project popularity in the OSS community. Notably, the numbers of followers and followees of an OSS project initiator have positive effects on the initiator's projects popularity. Moreover, the positive effect of the followee network also depends on the reach and importance of the initiator's followees. Specifically, the positive effect of following a new OSS user increases when the followee has high importance and reach. This way, our study not only enhances our understanding of the influencing factors for a strategically important outcome to OSS initiators (i.e., popularity), but also provides nuanced insights as to how to strategically follow to reap the most benefits of following.

References

Ahuja, G. 2000. "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study," *Administrative Science Quarterly* (45:3), Johnson School at Cornell University, pp. 425–455.

Borgatti, S., and Foster, P. 2003. "The Network Paradigm in Organizational Research: A Review and Typology," *Journal of Management* (29:6), pp. 991–1013.

Borgatti, S.P., and Halgin, D.S. 2011. "On Network Theory," *Organization Science* (22:5), pp. 1168–1181.

Borges, H., and Valente, M. 2018. "What's in a Github Star? Understanding Repository Starring Practices in a Social Coding Platform," *Journal of Systems and Software* (146), pp. 112–129.

Burt, R.S. 1992. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Chengalur-Smith, I., Sidorova, A., and Daniel, S. 2010. "Sustainability of Free/Libre Open Source Projects: A Longitudinal Study," *Journal of the Association for Information Systems* (11:11), pp. 657–683.

Chengalur-Smith, S., and Sidorova, A. 2003. "Survival of Open-Source Projects: A Population Ecology Perspective," in *24th International Conference on Information Systems*, Seattle, Washington.

Coleman, J.S. 1988. "Social Capital in the Creation of Human Capital," *American Journal of Sociology* (94), pp. S95–S120.

Coleman, J. S. 1990. *Foundations of Social Theory*, Cambridge, MA: Belknap Press.

Daniel, S., Agarwal, R., and Stewart, K. J. 2013. "The Effects of Diversity in Global, Distributed Collectives: A Study of Open Source Project Success," *Information Systems Research* (24:2), pp. 312–333.

Daniel, S., Midha, V., Bhattacherhjee, A., and Singh, S. 2018. "Sourcing Knowledge in Open Source Software Projects: The Impacts of Internal and External Social Capital on Project Success," *The Journal of Strategic Information Systems* (27:3), pp. 237–256.

Faraj, S., and Johnson, S. 2011. "Network Exchange Patterns in Online Communities," *Organization Science* (22:6), pp. 1464–1480.

Faraj, S., and Shimizu, T. 2018. "Online Communities and Knowledge Collaborations," *Oxford Research Encyclopedia of Business and Management*.

Github. 2017. "Saving Repositories with Stars." (https://help.github.com/articles/about-stars, accessed August 29, 2021).

Github. 2021. "Advanced Search." (https://github.com/search/advanced, accessed August 29, 2021).

Goes, P., Lin, M., and Yeung, C. A. 2014. "'Popularity Effect' in User-Generated Content: Evidence from Online Product Reviews," *Information Systems Research* (25:2), pp. 222–238.

Granovetter, M. 1983. "The Strength of Weak Ties: A Network Theory Revisited," *Sociological Theory* (1:1), pp.

201–233.
Granovetter, M. S. 1973. "The Strength of Weak Ties," *American Journal of Sociology* (78:6), pp. 1360–1380.
Grewal, R., Lilien, G., and Mallapragada, G. 2006. "Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems," *Management Science* (52:7), pp. 1043–1056.
Gulati, R., Puranam, P., and Tushman, M. 2012. "Meta-organization Design: Rethinking Design in Interorganizational and Community Contexts," *Strategic Management Journal* (33:6), pp. 571–586.
Hansen, M. T. 1999. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits," *Administrative Science Quarterly* (44:1), pp. 82–111.
Jarczyk, O., Gruszka, B., Jaroszewicz, S., Bukowski, L., and Wierzbicki, A. 2014. "Github Projects. Quality Analysis of Open-Source Software," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (8851), pp. 80–94.
Jiang, Q., Tan, C., Sia, C., and Wei, K. 2019. "Followership in an Open-Source Software Project and Its Significance in Code Reuse," *MIS Quarterly* (43:4), pp. 1303–1319.
Kadushin, C. 2012. *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford university press.
Kilduff, M., and Tsai, W. 2003. *Social Networks and Organizations*. Sage.
Krackhardt, D. 1992. "The Strength of Strong Ties," in *Networks and Organizations: Structure, Form and Action*, N. Nohria and R. G. Eccles (eds.), Boston, MA: Harvard Business School Press, pp. 216–239.
Kuk, G. 2006. "Strategic Interaction and Knowledge Sharing in the Kde Developer Mailing List," *Management Science* (52:7), pp. 1031–1042.
Lin, B., Robles, G., and Serebrenik, A. 2017. "Developer Turnover in Global, Industrial Open Source Projects: Insights from Applying Survival Analysis," in *2017 IEEE 12th International Conference on Global Software Engineering (ICGSE)*, pp. 66–75.
Long, Y. 2006. *Social Structure, Knowledge Sharing, and Project Performance in Open Source Software Development*, Lincoln, NE: University of Nebraska.
MarketResearchEngine. 2021. "Open Source Services Market." (https://www.marketresearchengine.com/open-source-services-market, accessed August 29, 2021).
Medappa, P., and Srivastava, S. 2019. "Does Superposition Influence the Success of FLOSS Projects? An Examination of Open-Source Software Development by Organizations and Individuals," *Information Systems Research* (30:3), pp. 764–786.
Moqri, M., Mei, X., Qiu, L., and Bandyopadhyay, S. 2018. "Effect of 'Following' on Contributions to Open Source Communities," *Journal of Management Information Systems* (35:4), pp. 1188–1217.
Morgan, L., and Finnegan, P. 2014. "Beyond Free Software: An Exploration of the Business Value of Strategic Open Source," *The Journal of Strategic Information Systems* (23:3), pp. 226–238.
Nahapiet, J., and Ghoshal, S. 1998. "Social Capital, Intellectual Capital, and the Organizational Advantage," *Academy of Management Review* (23:2), pp. 242–266.
Obstfeld, D. 2005. "Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation," *Administrative Science Quarterly* (50:1), pp. 100–130.
Ocasio, W. 1997. "Towards an Attention-Based View of the Firm," *Strategic Management Journal* (18:S1), pp. 187–206.
Peng, G., Wan, Y., and Woodlock, P. 2013. "Network Ties and the Success of Open Source Software Development," *The Journal of Strategic Information Systems* (22:4), pp. 269–281.
Setia, P., Bayus, B., and Rajagopalan, B. 2020. "The Takeoff of Open Source Software: A Signaling Perspective Based on Community Activities.," *MIS Quarterly* (44:3), pp. 1439–1458.
Setia, P., Rajagopalan, B., Sambamurthy, V., and Calantone, R. 2012. "How Peripheral Developers Contribute to Open-Source Software Development," *Information Systems Research* (23:1), INFORMS Inst.for Operations Res.and the Management Sciences, pp. 144–163.
Shapiro, C., and Varian, H. R. 1998. *Information Rules: A Strategic Guide to the Network Economy*, Boston, MA: Harvard Business Press.
Simon, H. 1971. "Designing Organizations for an Information-Rich World," in *Computers, Communications, and the Public Interest*, Baltimore, MD: Johns Hopkins University Press, pp. 37–52.
Singh, P., Tan, Y., and Mookerjee, V. 2011. "Network Effects: The Influence of Structural Capital on Open

Source Project Success," *MIS Quarterly* (35:4), pp. 813–829.
Stewart, K., Ammeter, A., and Maruping, L. 2006. "Impacts of License Choice and Organizational Sponsorship on User Interest and Development Activity in Open Source Software Projects," *Information Systems Research* (17:2), pp. 126–144.
Sutanto, J., Jiang, Q., and Tan, C. 2021. "The Contingent Role of Interproject Connectedness in Cultivating Open Source Software Projects," *The Journal of Strategic Information Systems* (30:1), p. 101598.
Sutanto, J., Kankanhalli, A., and Tan, B. 2014. "Uncovering the Relationship between OSS User Support Networks and OSS Popularity," *Decision Support Systems* (64), pp. 142–151.
Sykes, T. A., and Venkatesh, V. 2017. "Explaining Post-Implementation Employee System Use and Job Performance: Impacts of the Content and Source of Social Network Ties," *MIS Quarterly* (41:3), pp. 917–936.
Travers, J., and Milgram, S. 1969. "An Experimental Study of the Small World Problem," *Sociometry* (32:4), pp. 425–443.
Tsay, J., Dabbish, L., and Herbsleb, J. 2014. "Influence of Social and Technical Factors for Evaluating Contribution in GitHub," in *Proceedings of the 36th International Conference on Software Engineering*, , May 31, pp. 356–366.
Wang, X., Butler, B., and Ren, Y. 2013. "The Impact of Membership Overlap on Growth: An Ecological Competition View of Online Groups," *Organization Science* (24:2), pp. 414–431.
Wellman, B. 1997. "Structural Analysis: From Method and Metaphor to Theory and Substance," *Contemporary Studies in Sociology* (15), pp. 19–61.
Wen, W., Forman, C., and Graham, S. 2013. "Research Note—The Impact of Intellectual Property Rights Enforcement on Open Source Software Project Success," *Information Systems Research* (24:4), pp. 1131–1146.

Appendix A. Relevance of the 2012-2013 Dataset.

The purpose of this appendix is to provide the background of the dataset used in this study. We intend to demonstrate (1) the unique value of this dataset to our study, and (2) the relevance and validity of our findings using this dataset.

1. The unique value of this dataset to our study.

The 2012-2013 dataset is one of the most recent available datasets that provides an accurate depiction of the FollowEvent, due to changes in the data capture of the GitHub API that happened in December 2013. A newer dataset will not have accurate information about the FollowEvent. We here provide supportive evidence.

Moqri et al. (2018) indicate that "*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.*" (p. 1199)

To validate and understand this statement, we use ghtorrent.org – a GitHub archive repository –to confirm that the number of followers suddenly dropped in December 2013. Figure A1 illustrates the sudden change in the number of followers.

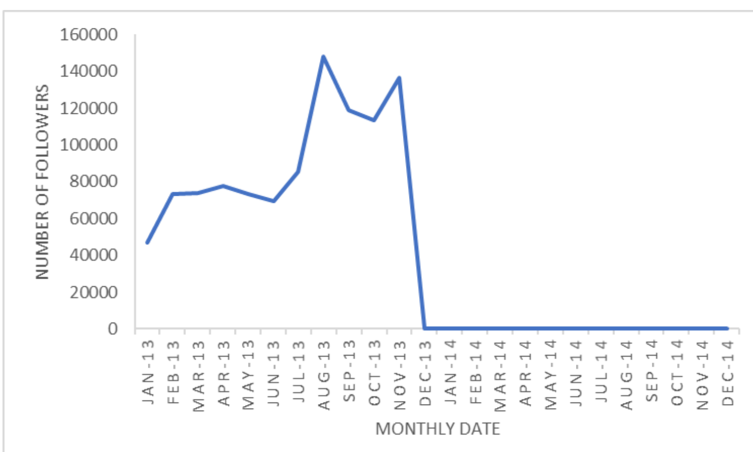


Figure A1. GitHub Followers After January 2013

Considering that GHTorrent collects data from the GitHub REST API (see Table 1 of Mombach and Valente 2018, shown below), we realize that this change is caused by GitHub not providing the API for the FollowEvent. We find several references that confirm that GitHub API stopped providing the FollowEvent (see. Table A1).

	GitHub REST API V3	GitHub Archive	GHTorrent
Data Since	-	02/12/2011	2012
Data From	-	GitHub REST API Events	GitHub REST API
Data Type	GitHub Data	GitHub Events	Structured GitHub Data
Data From	Live	Hourly	Monthly

Table 1. Analysis of GitHub REST API V3, GitHub Archive, and GHTorrent.

Table A1. GitHub API Stops Providing FollowEvent

FollowEvent Triggered when a user follows another user. Events of this type are **no longer created**, but it's possible that they exist in timelines of some users. Events API payload Key / Type / Description: `target` / `object` / The user that was just followed. Webhook event name `follow`	FollowEvent Triggered when a user follows another user. Events of this type are **no longer created**, but it's possible that they exist in timelines of some users. Event name `follow` Payload Key / Type / Description: `target` / `object` / The user that was just followed.
github.com/github/developer.github.com/blob/master/content/v3/activity/events/types.md	docs2.lfe.io/v3/activity/events/types/#followevent

In conclusion, 2012-2013 is a unique dataset that enables us to more accurately explore FollowEvents. Because GitHub does not provide the API for FollowEvent, we follow Moqri et al. (2018) by limiting the dataset to 2012-2013.

At the same time, we admit that the StarEvent (also called the WatchEvent) is not unique to our dataset. That means newer datasets (i.e., after December 2013) still contain the StarEvent. Below is the description from GitHub (//github.com/github/developer.github.com/blob/master/content/v3/activity/events/types.md).

WatchEvent

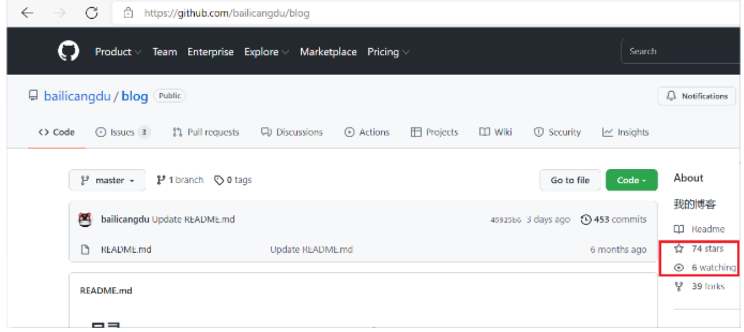
The WatchEvent is related to starring a repository, not watching. See this API blog post for an explanation.

The event's actor is the user who starred a repository, and the event's repository is the repository that was starred.

In Table A2, on the left, we show the API results, and on the right, we present the display on the platform. The WatchEvent returns the number of stars.

Table A2. GitHub WatchEvent

api.github.com/users/bailicangdu/repos



github.com/bailicangdu/blog

The GitHub API documentation states: "*We recently changed the Watcher behavior on GitHub. What used to be known as 'Watching' is now 'Starring'. Starring is basically a way to bookmark interesting repositories. Watching is a way to indicate that you want to receive email or web notifications on a Repository.*" (developer.github.com/changes/2012-09-05-watcher-api/).

After the change in August/September 2012, the WatchEvent kept returning the number of stars. GitHub introduced the *subscribers'* endpoint to return watchers (docs.github.com/en/rest/reference/activity#watching; docs.github.com/en/rest/reference/activity#starring).

Therefore, a newer dataset will still have accurate information about the StarEvent but will not have accurate information about the FollowEvent. GHTorrent states that "the *created_at* field is only filled in accurately for followships for which GHTorrent has recorded a corresponding event. Otherwise, it is filled in with the latest date that the corresponding user or follower has been created." (ghtorrent.org/relational.html). Because the FollowEvent is stopped after December 2013, the exact date of the creation of the FollowEvent may not be accurate. The additional dataset (January 2016 – January 2017) presents this limitation, and although the results corroborate our main findings, we believe that the results should be interpreted in light of the current limitations of the dataset.

Because of the limitations we mentioned above (i.e., lack of exact date of creation of the FollowEvent), it may not be a great idea to collect a dataset newer than 2017. GHTorrent provides links to download data dumps (ghtorrent.org/downloads.html). The main reason why we choose to download the January-2017 data dump is because of our current processing limitations. A large data dump requires a longer time and more computing power to process. Although the Zip file mysql-2017-02-01 is about 50146 MB, the extracted data size is already nearly 100 GB (see Figure A2).

- mysql-2017-10-01 (62416 MB)
- mysql-2017-09-01 (61281 MB)
- mysql-2017-07-01 (58125 MB)
- mysql-2017-06-01 (56567 MB)
- mysql-2017-05-01 (55047 MB)
- mysql-2017-04-01 (53516 MB)
- mysql-2017-03-01 (50465 MB)
- mysql-2017-02-01 (50146 MB)
- mysql-2017-01-19 (49553 MB)
- mysql-2016-09-05 (43186 MB)
- mysql-2016-07-19 (41318 MB)
- mysql-2016-06-16 (40457 MB)
- mysql-2016-06-01 (39851 MB)
- mysql-2016-05-04 (38618 MB)
- mysql-2016-04-19 (38247 MB)
- mysql-2016-03-16 (36914 MB)
- mysql-2016-03-01 (36228 MB)
- mysql-2016-02-16 (35574 MB)
- mysql-2016-02-01 (34969 MB)

Figure A2. GHTorrent Data Dumps

(2) The relevance and validity of our findings using this dataset.

We further research and examine the major updates and changes by GitHub to confirm that none of them will influence our findings. GitHub keeps an archive on changes on both their API and platform. In Table A3, we illustrate some of the changes on the API & platform.

Table A3. Examples of Changes on the API and Platform

	API resource changes (API user) GitHub API Changes | GitHub Developer Guide The GitHub Blog | Updates, ideas, and inspiration from GitHub to help developers build and design software.	Platform changes (end-user) The GitHub Blog | Updates, ideas, and inspiration from GitHub to help developers build and design software.
October 2012	Public member list Organization Members Resource Changes | GitHub Developer Guide	Username auto-completion GitHub for Mac: Username Autocompletion | The GitHub Blog
	Rate limits Rate limit changes for unauthenticated requests | GitHub Developer Guide	Notifications GitHub for Mac: Notifications | The GitHub Blog New close and merge notifications | The GitHub Blog
	Gist comments Gist comment URIs | GitHub Developer Guide	Emoji autocomplete Emoji autocomplete | The GitHub Blog

	Notifications Notifications API | GitHub Developer Guide	Latest commit Latest commit per directory | The GitHub Blog
	Repository Set the default branch for a repository | GitHub Developer Guide	
September 2012	Add commits to a repository Initialize a repository when creating | GitHub Developer Guide	Command bar Introducing the Command Bar | The GitHub Blog
	Changes to Watches vs Stars Upcoming Changes to Watcher and Star APIs | GitHub Developer Guide	Searching and filtering stars Searching and Filtering Stars | The GitHub Blog
		New launch page GitHub Launch Page | The GitHub Blog
		New user profile page New User Profile Pages | The GitHub Blog
December 2012	Introduction of diff and patch Diff and patch media types | GitHub Developer Guide	Pull requests Tidying up after Pull Requests | The GitHub Blog
	Pagination Pagination for Organization Repository lists now paginates properly | GitHub Developer Guide	Status site New Status Site | The GitHub Blog
	Repository Finding sources and fork repositories for organizations | GitHub Developer Guide	Files creation Creating files on GitHub | The GitHub Blog
	OAuth Create an OAuth authorization for an app | GitHub Developer Guide	Issues attachment Issue attachments | The GitHub Blog
	Review and Issue comment Per-repository Review and Issue Comment listing | GitHub Developer Guide	Gists Welcome to a New Gist | The GitHub Blog
	Status API GitHub system status API | The GitHub Blog	Uploads Goodbye, Uploads | The GitHub Blog
		Homepage New Homepage | The GitHub Blog
		Receipts Customize your receipt details | The GitHub Blog
		Issues autocompletion Issue autocompletion | The GitHub Blog
November 2012	Repository Gitignore Templates API | GitHub Developer Guide	Autocomplete stars Command bar autocompletes stars | The GitHub Blog
	Forking Forking to Organizations | GitHub Developer Guide	New header & footer New header and footer | The GitHub Blog
		Impact graphs Retiring Impact Graphs | The GitHub Blog
		Search syntax Search Syntax Improvements | The GitHub Blog

To strengthen the validity of our findings, we identify the changes related to the Follow and Star activities (see Table A4). We focus on these two activities because they are the basis of our hypotheses. Based on our review of the changes, we can conclude that there are no changes that fundamentally affect the Follow and Star activities. Most of the changes build on Follow and Star activities to provide additional features (e.g., sort repositories based on stars). Additionally, we interview a "veteran" GitHub user with 10-year use experience, who confirms that no major functionality or structural changes to GitHub are noticed that would affect followership and starring.

Table A4. Changes on the API and Platforms: Follow and Star Activities

	API resource changes (API user) GitHub API Changes | GitHub Developer Guide The GitHub Blog | Updates, ideas, and inspiration from GitHub to help developers build and design software.	Platform changes (end-user) The GitHub Blog | Updates, ideas, and inspiration from GitHub to help developers build and design software.
	Star	
June 2012		Starring Gists Starring Gists | The GitHub Blog
August 2012		Stars & notifications Notifications & Stars | The GitHub Blog
September 2012	Stars & watches Upcoming Changes to Watcher and Star APIs | GitHub Developer Guide	Searching and filtering stars Searching and Filtering Stars | The GitHub Blog Watcher behavior Watcher API Changes | The GitHub Blog
November 2012		Autocomplete stars Command bar autocompletes stars | The GitHub Blog
February 2013	Sort stars Sortable Stars in Repository Starring API | GitHub Developer Guide	Sort stars Sortable Stars | The GitHub Blog
July 2013		Explore stars Explore Everyone's Stars | The GitHub Blog
December 2013		Explore More Explore Features | The GitHub Blog
February 2014		People you know People you know | The GitHub Blog
December 2014	New attributes for Starring New Attributes for Starring API | GitHub Developer Guide	
January 2019		Starring topics Star topics that interest you | The GitHub Blog
April 2019	Webhooks New webhook events and actions | GitHub Developer Guide	
April 2020		Stars program Introducing the GitHub Stars Program ⭐ | The GitHub Blog

April 2021		Sort repositories by stars Sort repos in user and organization profiles by star count | GitHub Changelog
December 2021		Lists Lists are now available as a public beta | GitHub Changelog
Follow		
April 2008		Follow Who are you following? | The GitHub Blog
June 2011	API New Users, Repos, and Orgs API | The GitHub Blog	
January 2013	User scope New User scopes | GitHub Developer Guide	
December 2013		More explore features More Explore Features | The GitHub Blog
February 2014		People you know People you know | The GitHub Blog
August 2017	Webhooks Breaking changes to creating webhooks via the API | GitHub Developer Guide	
October 2018	Additional endpoints Additional endpoints available for GitHub Apps | GitHub Developer Guide	

References.

Mombach, T., & Valente, M. T. 2018. "GitHub REST API vs GHTorrent vs GitHub Archive: A comparative study".

Moqri, M., Mei, X., Qiu, L., and Bandyopadhyay, S. 2018. "Effect of 'Following' on Contributions to Open Source Communities," *Journal of Management Information Systems* (35:4), pp. 1188–1217.