The Dissertation Committee for Wei Yang Certifies that this is the approved version of the following Dissertation:

TECHNOLOGY ENTREPRENEURSHIP AND VALUE CREATION ON OPEN INNOVATION PLATFORMS

Committee:

Francisco Polidoro Jr., Supervisor

Puay Khoon Toh

Ramkumar Ranganathan

Wen Wen

TECHNOLOGY ENTREPRENEURSHIP AND VALUE CREATION ON OPEN INNOVATION PLATFORMS

by

Wei Yang

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

2019

Abstract

Technology Entrepreneurship and Value Creation on Open Innovation Platforms

Wei Yang, PhD

The University of Texas at Austin, 2019

Supervisor: Francisco Polidoro Jr.

This dissertation studies how entrepreneurial firms create economic value from open source technology platforms, interfaces on which firms disclose knowledge and distribute innovation for free without retaining any proprietary rights. Despite their increasing importance in innovation and growing popularity among profit-seeking new ventures, open source platforms present a major challenge for value creation, as they lack price signals to guide ventures' transactions and forfeit ventures' control over key resources and knowledge for innovation. Those features are in contrast with the fundamental assumption about price and revenue in economics. They also run counter to the central tenet in strategy research that private knowledge and rare resources are central to competitive advantage and profiting from innovation.

To address this puzzle about value creation from free technologies base on free knowledge and resources, this dissertation specifically focuses on the economic implications of strategies ventures can leverage within and across open source development communities. Chapter I reviews the literature relevant to entrepreneurship in an open and inter-dependent innovation environment. Exploring research opportunities emerged from the literature review, Chapter II explores the possibility that multihoming, a critical growth strategy of ventures as open source complementors in platform competition, allows ventures to reinforce their existing user base – a prerequisite of

value creation from open source. Chapter III directly addresses value creation by investigating how collaborating with external contributors, another critical open source strategy, influences venture capital investment. Both essays highlight how platform network effects unfold without price signals and proprietary rights of the technologies in shaping the outcome for ventures' strategies. They also emphasize those strategies' demand side implications on users, participants on another side of open source platforms.

The empirical analyses of this dissertation are based on multiple open source technologies platforms, with data obtained from on GitHub, the worlds' largest open source software storage provider, containing 5 Terabytes of information on 2.1 million ventures, 96 million technologies and over 2 billion development activities, under research designs for deriving causal references. Overall, the dissertation seeks to advance the understanding of value creation in entrepreneurship through open source platforms, an increasingly important phenomenon in contemporary economy.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Protecting intellectual property rights is central for firms to profit from technological innovation. The control of knowledge through well-defined intellectual property rights in tight appropriately regime allows firms to effectively retain rare and valuable knowledge resources, deter imitation, mitigate transactions costs, and reduce hazards of misappropriation, all of which are critical to the creation and capture of value in technology-intensive settings (Barney, 1991; Coase, 1960; Cohen, Nelson, & Walsh, 2000; Gulati & Singh, 1998; Oxley, 1997; Peteraf, 1993; Teece, 1986; Williamson, 1985). Protecting firm knowledge through intellectual property strategies such as patenting is particularly essential for new ventures, as they oftentimes lack the bargaining power and complementary assets and capabilities to compete in the downstream product markets (Gans & Stern, 2003; Pisano, 1990).

However, this long-established view regarding the control of knowledge and value creation has been increasingly challenged by the growing popularity of open source innovation among profit-seeking new ventures (Alexy, West, Klapper, & Reitzig, 2018; Colombo, Piva, & Rossi-Lamastra, 2014; Fosfuri, Giarratana, & Luzzi, 2008; Wen, Ceccagnoli, & Forman, 2015). Different from the proprietary innovation process where firms strategies and competition center on the protection of critical knowledge against misappropriation, in open source, firms not only develop and distribute technologies for free, but also allow public access to all the underlying technical details and knowledge, in a way that any external parties can modify and redistribute the innovation to anyone and for any purpose (Levine & Prietula, 2013; Von Krogh & Von Hippel, 2006).

Although initially emerged as a movement against commercial innovation (Bonaccorsi & Rossi, 2003; Von Krogh & Von Hippel, 2003), open source has become increasingly relevant to profit-seeking firms in many high technology industries. On the one hand, open source is gaining increasing technological importance, as it breeds considerable path-breaking technologies that are disrupting the existing proprietary technologies with their unprecedented impact on the economy and society (Tucci, Afuah, & Viscusi, 2018)[1]. The cumulative innovation among open source technologies, unbounded by intellectual property rights, has enabled the overall knowledge creation to grow at an exponential rate, which makes the practice also appealing to resource constraint values who constant search for valuable knowledge inputs (Nagle, 2018). While earlier literature contends that firms tend to regard open source as a threat and resort to commercial proprietary innovation to compete against such technologies (Bonaccorsi & Rossi, 2003; Economides & Katsamakas, 2006; Von Krogh & Von Hippel, 2003), in recent years, firms, especially resource constraint new ventures, are increasingly prone to participate in open source due to such technological impacts and knowledge benefits it demonstrates (Alexy & Reitzig, 2013; Alexy et al., 2018).

On the other hand, new ventures are increasingly attracted to open source, because it has revealed the huge business opportunities and potential of economic value creation potential for entrepreneurship. Although built on public knowledge and distributed for free, open source technologies have created over $147 billion economic value in entrepreneurship, with nine IPOs currently valuing at $67 billion, over 200 mergers and acquisitions that involved over $20 billion, and over 10 thousands rounds of venture capital investment that involved $10 billion to their developing ventures, which have given rise to over 40 ventures with 100 million valuation and multiple unicorns that exceeds a billion valuation in U.S. dollars in the past decade (Jacks, 2018;

[1] Those path-breaking innovations, technologies, while largely within the field of computer science, ranges from mobile communication, block-chain transactions or automated driving and aviation, to technologies that defeated the smartest chess players of the world, or allow problem-involving based on massive amounts of data and computation, all of which have profoundly influenced the contemporary innovation and overall development of the society.

Myers, 2018; Rowley, 2017). The economic prospect of open source technologies stimulates active entrepreneurship in open source communities (Wen et al., 2015).

The growing popularity and huge economic value of open source present an intriguing puzzle yet to be fully addressed in strategy research. As mentioned at the beginning, open source entrepreneurship without control of knowledge and resources runs counter to the central tenet in strategy research that private knowledge and appropriability regime is key to competitive advantage (e.g., Barney, 1991; Teece, 1986; Wernerfelt, 1984). Moreover, due to the unique presence of developer/user communities in open source innovation, ventures tend to rely more heavily on the external knowledge and inputs from to free external contributors for knowledge creation (Nagle, 2018), which is also in contrast with what we know about the role of knowledge in explaining the very existence of firms (Grant, 1996; Kogut & Zander, 1992; Nickerson & Zenger, 2004). Moreover, as open source technologies are distributed for free, the lack of price mechanisms on the corresponding markets makes the value creation more difficult to comprehend given that given the fundamental role of price in profit commonly assumed in economics. Also, entrepreneurship through open source has demonstrated even greater variance in performance and survival. While a few, as discussed earlier, gained substantial technological and economic success, most open source technologies fail to attract any market attention, even though they are made free with all underlying knowledge disclosed – a drastic heterogeneity that is rarely discussed. Those tensions give rise to the research question of this dissertation - *how can firms, especially new ventures gain competitive advantage and profit from developing open source technologies that are distributed for free and without proprietary rights of their knowledge, on markets with intense competition and lack price signals?*

To explore this research question about the value creation from open source technologies, this dissertation specifically investigates the strategies ventures leverage within and across the development communities during open source innovation, and their impacts on the key outcomes of value creation in entrepreneurship. The focus of development communities in value creation first originates from their importance in open source innovation. Different from conventional innovation, the development of open source technologies mostly happens in the communities, through the constant interactions between the sponsoring organization of the key technological infrastructure and a variety of participants, including users, external contributors and suppliers of complementary technologies and the venture (e.g., Foss, Frederiksen, & Rullani, 2016; O'Mahony & Ferraro, 2007; Von Krogh, Spaeth, & Lakhani, 2003). Such unique innovation process makes the strategies that can shape participants' behavior within communities particularly relevant to the economic value of the open source technologies.

In discussing the value creation of open source innovation, however, current research has not yet explored the implications of strategies and dynamics within those communities, with the current emphasis on business models and competition with proprietary innovation (e.g., Massa, Tucci, & Afuah, 2017; Teece, 2007). The discussion on open source value creation is disconnected from the unique community-based innovation process of open source technologies. At the same time, while other studies on open source communities have explored innovation process in open source technologies, such discussion ignores the value creation possibilities of those communities by emphasizing the voluntary and anti-commercial nature of open source communities and all the participants involved (e.g., Hertel, Niedner, & Herrmann, 2003; Roberts, Hann, & Slaughter, 2006) – an increasingly questionable assumption given the active entrepreneurship and incumbent tech-giants in open source communities (Asay, 2016; Silver, 2018).

Moreover, this underlying assumption of the nonprofit nature of open source technology communities has directed much of the attention to the competition between open source and proprietary technologies. In turn, we lack the understanding of the competitive dynamics within open source technologies and communities. Few studies have addressed the huge variance of success across open source technologies and their communities – that is, why some technologies outcompete others, given they are equally distributed without price and knowledge protection and have the same access to the public who can potentially contribute to the subsequent innovation of the technologies, and how is such heterogeneity related to the strategies of ventures in the open source communities?

Addressing those tensions around the value creation of open source and the influence of ventures' strategic behavior in the corresponding technology communities, the theory development of this dissertation conceptualizes open source communities as multi-sided technology platforms, with the competition among open source technologies as a platform-based process without price signals. On such platforms, the interactions among users, contributors and new ventures are mediated by a same set of open source technology infrastructures. Ventures can create and capture value by resuming a variety of roles, either by becoming sponsors (owners) of the communities or becoming complementors that supplies add-on technologies and knowledge. By regarding open source communities as multi-sided technology platforms, this dissertation highlights the role of network effects, unique to such environment, in shaping the economic outcome of ventures' strategies. At the same time, different from the existing literature that focuses on how open source community platforms supply and create knowledge inputs on the upstream (Belenzon & Schankerman, 2015; Stam, 2009), this dissertation highlights the influence of community platforms in the downstream market competition for users, who constitute critical resources for direct value creation. In doing so, this dissertation connects community platforms with value creation of open source.

More specifically, this dissertation investigates the value creation implications of two critical strategies ventures can leverage on open source community platforms, each corresponding to the two types of roles ventures can take on those platforms, as complementary technology providers (complementors), or as initiators/owners (sponsors) of the community platforms. Chapter I starts with a comprehensive literature review on the current state of research relevant to entrepreneurship based on open innovation without proprietary rights, followed by the identification of research opportunities arising from those pockets of literature regarding the boundary decision made by ventures in such environment. The review of three most relevant pockets of literature namely (1) open innovation and open source technologies (2) multi-sided technology and product platforms (3) technologies ecosystems show that existing literature has yet fully addressed how and why growth and performance of new ventures vary in a competitive environment without price signal, while accounting for the high technological interdependency due to the public nature of innovation and knowledge.

Following the research opportunities identified from the literature review, the first empirical essay in Chapter II focuses on the expansion strategies of ventures as complementors on open source platforms. The essay explores the technological consequences of new venture growth in open source platforms by investigating how a venture’s expansion to multiple open platforms (referred as multihoming) affects its existing user base, a critical prerequisite of value creation through virtually all business model of open source. Due to the cumulative nature of open innovation, technologies are usually platforms based, in a way that creates considerable entrepreneurship opportunities for ventures as complementors to major open source technologies

(Economides & Katsamakas, 2006; West, 2003). Prior research has extensively examined the performance implications of the broadening of a firm's scope across industries (e.g., Chatterjee & Wernerfelt, 1991; Krishnan, Miller, & Judge, 1997; Miller, 2006; Montgomery & Wernerfelt, 1988). Yet, research is yet to examine whether existing insights apply to open innovation platforms, in which most providers of complementary products are entrepreneurs and small ventures. Unlike incumbent firms possessing slack resources, they are resource-constrained and with the limited protection of intellectual property rights in open source. Strategy research on platforms has highlighted the performance consequences of technological interdependencies within a platform but has stopped short of investigating dynamics that unfold across platforms (Kapoor & Agarwal, 2017). The study proposes that a complementor's expansion to an alternative open innovation platform, a strategy referred to as multihoming, has a positive effect on its user base in the original platform. The theoretical development of this chapter highlights the transfer of platform network externalities for complementors through multihoming as the mechanism underlying the positive effect. More specifically, users prefer those multihoming complements as they allow boarder scope of interaction (direct inter-platform network effects) while lowering the learning cost of additional adoption on other platforms (indirect inter-platform network effects). Multihoming's positive effect on user base is also related to from the absence of prices signals on open source platforms. Expansion, then, signals the technological stability and certainty in a way that increases user's confidence in the technologies on the original platform.

The empirical analysis is based on data on 2 million software technologies in 34 open source software development platforms, under a matching design between multihoming ventures with similar counterfactual ventures that focus on providing complementary technologies to a single open source platform. The results provide strong support to the proposed hypotheses, while further showing that while user awareness strengthens the positive effect of multihoming, high technological interdependencies with other technologies and competitive advantage of the original open source platform tend to weaken the effect of multihoming in ventures' growth of user base.

Chapter III, then, shifts the focus to the direct value creation of open source by investigating the venture capital investment made to open source-based ventures. Compared with Chapter II that conceptualizes open source development platforms as two-sided markets, where the common technological infrastructure connected ventures as complementors with users, Chapter III extends the conceptualization of open source communities as multi-sided platforms, with the emphasis of the role another critical actor – the crowd as external contributors on the development communities as platforms. It examines the impacts of collaborating with the crowd, who are fundamental to open source community platforms, on the value creation of ventures sponsoring those platforms. In contrast with the existing literature that highlights the crowd as knowledge inputs, this study highlights the role of the crowd in providing ventures with access to critical market resources. Through the platform-based interaction and communication, the collaboration process familiarizes the crowd with ventures' innovation in terms of both knowledge and trusts, which create path-dependencies that lock in those external contributors. Furthermore, because the crowd oftentimes composes of lead users, they are also critical in attracting other ordinary users because of their prominent role in generating direct network effects (Lee& Lee 2006), facilitating the technology diffusion on the product market. The value of such market resources established through the crowd will affect venture capital investment, as the major reflection of ventures' economic value. Based on those mechanisms, the study also proposes that the positive effect of crowd collaboration will be weakened by the amount of knowledge venture disclosed to attract collaboration due to the increasing opportunity cost of making such knowledge non-proprietary. The positive effect of crowd collaboration is also accentuated by the diversity of the ventures' knowledge base, which allows ventures to attract different types of lead users while minimizing the overlap of ordinary users in the crowds' network effects. The hypotheses are fully supported in the empirical analysis, based on data from GitHub, the world's largest open source technology storage hosts, with information on 450,097 open source-based ventures, 14,472,957 records of collaboration and 10, 742 rounds of venture capital investment from 2013 to 2017.

This dissertation seeks to advance the understanding entrepreneurship based open source innovation from the following perspectives. First, this study directly connects the value creation from open source with the community dynamics and platform strategies of entrepreneurial firms. In contrast with the existing literature that focuses on business model innovation in studying how ventures profit from open source technologies, this dissertation directly explores how the value creation and value capture implications of the strategic behaviors of ventures in the course of such platform based technological innovation, especially with regard to growth and collaboration in platform-based competition. The focus of user base and market resources in such process discussed in the two empirical essays in Chapter II and Chapter III also highlights the demand-side dynamics triggered by key open source strategies, which is rarely explicitly discussed in current research.

Secondly, this study advances the understanding of the heterogeneity and competition among open source technologies. In contrast with most of the current research that focuses on one or a few communities, this demonstrates huge heterogeneity across community platforms and ventures, in terms of both the strategic behavior and related outcome. The highlight of the strategic choices of ventures, including growth and collaboration, seeks to address the origins of the heterogeneity in the value creation among competition open source technologies. In doing so, it also helps address the question that is not fully addressed in the current literature, that is, why only a few open source technologies are able to succeed, while the majority fail even if ventures fully disclose the knowledge and provide them for free?

Thirdly, as this dissertation conceptualizes open source innovation and competition as a platform-based process, it also contributes to the literature on platforms and two-sided markets. On the one hand, the investigation of open source as platforms that provide free technologies shed lights on the platform dynamics without price signals, a critical assumption and focus in the existing literature. On the other hand, the theorization about the transfer of network effects beyond the boundaries of competing platforms provides new insights in the understanding of platform competition. In addition, the focus of complementors' strategies also shifts the focus from the platform owners in the current literature to the strategies and implications of other critical actors on platforms in the course of platform competition, opening up new research possibilities to investigate complementor strategies on open platforms in the future.

This dissertation also bears important empirical contributions. The empirical analyses on the essays are based on unprecedentedly large data detailed to activity level, with Terabytes of information, which not only allows detailed measures of ventures' strategic behavior and outcome, but also the observation of heterogeneities among multiple platforms. On the one hand, the rich information in the data allows detections of variances the existing literature investigating one or a few communities has not captured. The massive data also allows identification of potential counterfactuals, which is the key to deriving causal inferences. On the other hand, this dissertation's use of unconventional big data to study research questions relevant to strategy and innovation also is among the first efforts that methodologically connect strategy research with the leading data and computation techniques, provides new insights that open up considerable opportunities for strategic research in the era of big data.

Chapter I. Literature Review: Current Research on Open Source Innovation

In this session, I review and critique the current literature that is essential to the theoretical and phenomenological focus of my dissertation. As the dissertation is essentially interested in studying the boundary and performance of high technology ventures when knowledge is open and interconnected, the literature review focuses on the current findings of innovation and competition dynamics in an environment where knowledge system underlying innovation could transcend firm boundaries. More specifically, I will summarize the state of research on the following four topics: (1) open source innovation, which focuses specifically on innovation and technology development based on public and non- proprietary knowledge from the supply side of innovation (2) platform-based competition, which highlights the externality of technology adoption and diffusion from the demand side (3) technology ecosystems, which emphasis on the interdependence of technology and innovation that transcend firm boundaries as a system. For each topic, I also identify the opportunities emerge from those studies, and elucidate they are connected to the potential contribution of my dissertation.

To search the relevant literature in those areas, I focused on the top journals in strategic management and entrepreneurship research, including (1) Academy of Management Review (2) Academy of Management Journal (3) Strategic Management Journal (4) Organization Science (5) Management Science (6) Research Policy. Because the platform-based research is partly rooted in economics, I also included several top economics outlets including (1) American Economic Review (2) RAND Journal of Economics (3) Journal of Economics and Management Strategy. While I didn't limit the year of publication in the search, the majority of the relevant papers are published after the 2000s. The search yielded 45 papers for open source innovation, 53 paper platform-based competition and 15 for platform ecosystem. The following literature review is largely based on those papers identified in the top journals.

OPEN SOURCE INNOVATION

Definition and overview

While open innovation is sometimes broadly used to describing inter-firm collaboration for innovation (e.g., Boudreau, 2010; Chesbrough, 2003; Christensen, Olesen, & Kjær, 2005), this dissertation focus on the most strictly defined form of open innovation that poses challenges to the boundary decisions unexplained in currently literature (e.g., Afuah & Tucci, 2013; Lichtenthaler, 2011). More specifically, *this dissertation refers open innovation as non-priced and non-proprietary technologies with underlying knowledge shared and distributed to the public* (Kogut & Metiu, 2001; Lerner & Tirole, 2005a; Von Krogh & Von Hippel, 2003). In other words, I focus on entrepreneurship based on open innovation in the form of open source, a model of innovation that is formally defined as "a decentralized…development model that encourages open collaboration…with products such as source code, blueprints, and documentation freely available to the public" (Wikipedia, 2018b)[2].

The concept of open source as the most open form of innovation was initially developed as an opposition to commercialized innovations in the context of software technology development (Von Krogh & Von Hippel, 2003). In the 1980s, open source emerged as a "movement" led by university scientists in computer science, as a protest to the university's decision that allowed a company to incorporate their computer codes in commercial software and profit from their knowledge (Von Krogh & Von Hippel, 2003). Among them, Richard Stallman founded Free Software Foundation, sought to institutionalize the open source practice through open source license. The General Public License (GPL) he developed entails "those possessing a copy of free software…the right to use it at no cost, the right to study its "source code," to modify it, and to

[2] In this dissertation, open innovation and open source can be regarded as interchangeable.

distribute modified or unmodified versions to others at no cost." (Von Krogh & Von Hippel, 2003: 1151: 1151). Such basic spirit has become the foundation of current open source innovation practice (Levine & Prietula, 2013).

Early research on open source from the perspective of innovation and technology evolution started in the late 1990s. The discussion open innovation all started with the phenomenon itself gaining importance, with early authors sought to describe it, and makes sense out of the phenomenon from a technological point of view. The earliest influential inquiry in open source is perhaps the book "The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary" by Eric Raymond (1998). Observing the huge success of open source software such as Linux and noting that "code for sale is just a tip of the iceberg" (Chapter 4). In the book, he first brought up the concept of open source as a "movement", as opposed to a mode of innovation. Summarizing the history of open source (up to 1998), he concluded that the rise of open source was driven by individual developers and profoundly rooted a strong sense of "Hecker culture" (Raymond, 1998: 2) that has a strong orientation to problem-solving, sharing and creativity, as opposed to profit maximization and commercialization. At the same time, Raymond also noted several critical issues emerged from the open source, including how ownership is defined without proprietary rights, causes of conflict in decision making, the nature of the contribution and open source community. While the book was largely descriptive, without elaborating mechanisms driving such increasingly important phenomenon, the characterization of the open source innovation as a unique "culture" and "movement", as well as the critical issues about open source it raised, deeply shaped how open source was regarded and researched in the later works.

Arguably the first attempt to address open innovation from an academic viewpoint to gain systematize understanding is the editorial of Research Policy in 2003. In 2003, Research Policy launched a special issue on open source software development, almost 20 years after the start of open source. In the editorial, Von Krogh and Von Hippel (2003) summarized three critical research directions of open source at the time, namely, (1) motivation of contributors (2) innovation process, which focuses on the governance and growth of open source community (3) "competitive dynamics", which highlights how to reconcile open source with proprietary and commercial technologies (Von Krogh and Von Hippel, 2003: 1152-1155). Such distinction is further reinforced in a special issue on open source in Management Science three years later in 2006. The first perspective, the motivation of contributors, explores the formation of external resource supply on open innovation environment without price and clearly defined ownership. Such external supply of human resource capital is a critical alternative to ventures' internal resource accumulation in open innovation. The second perspective, the innovation process, largely focuses on the how labor supply external to ventures is governed and how technologies are developed based on open knowledge. The third perspective, the dynamics competitive, emphasizes the performance implication of open innovation. This structure gives a clear roadmap on the issues in open innovation relevant to this dissertation. Hence, in the literature review on open innovation, I will largely follow this established template and then summarize the characteristics of open source in current research and discuss important issues that have yet been addressed in the current literature, laying down the fundamental motivation of my dissertation.

Motivation of contributors

Early research on open innovation and open source displayed considerable interest in exploring the motivation of contributors to open source technologies. Different from the traditional knowledge labor on the market or within firm boundaries, who work in exchange of salary or ownership of the intellectual property from which they can profit, many developers voluntarily develop and improve open source software without financial compensation and to not retain intellectual property. Such phenomenon in open source prompted scholar to explore the question, "Why should thousands of top-notch programmers contribute freely to the provision of a public good?" (Hippel & Krogh, 2003: 212).

To date, the extensive literature on the contributor motivation and the resulting behavior have provided detailed answers. Those answers can be categorized into four categories (1) intrinsic motivation (2) extrinsic motivation (3) technological characteristics, based on the focus of factors underlying contributor's behavior.

2.1. Intrinsic motivation

The first perspective in open source research explores the intrinsic motivation of contributor behavior (Bagozzi & Dholakia, 2006; Belenzon & Schankerman, 2015; Hertel et al., 2003; Krishnamurthy, Ou, & Tripathi, 2014). Compared with knowledge workers for corporations, innovators in open source are driven by fundamentally different ideologies towards innovation (Von Krogh et al., 2003). Developers are willing to contribute and innovate for free, because they believe that knowledge should be public goods rather than a tool for profit (Krishnamurthy et al., 2014). Bagozzi and Dholakia (2006), for example, demonstrated how the cognitive and affective factor related to contributor's perception towards open source play a critical role in determining the extent to which developers participate in the development of Linux kernel based on a survey of 191 developers worldwide. Theorizing open source contribution as group-referent intentional actions, they predict and found that developer's positive attitudes, emotions and social identifies all positive influenced their tendencies to participate the Linux development, because those factors influence the decision making, as well as developers' perception of self-worth when joining and contributing to open source innovation. In a qualitative study of 70 open source contributors, O'Mahony (2003) found that the intent to keep their knowledge open and part of the communities contributes significantly to the contributing behavior of participants, further revealing that the important role of developers' ideological belief in the nature of knowledge is an important factor underlying such behavior. Franke and Von Hippel (2003), on the other hand, understood contributors' intrinsic motivation in a slightly different way. Through the case study of Apache foundation, another extremely successful open source initiative, they argue that a critical motivation for open source contribution is developer's own needs of using the innovation. Because the demand of users are highly heterogeneous, to an extent that it is difficult for a single innovator to satisfy, users are motivated to participate and contribute to the innovation, so that they can better utilize the technologies for their own purposes. Shah (2006) further distinguished two types of intrinsic motivation as "hobbyist" and "need driven". Through inductive studies of two open source communities, she found that hobbyist has a longer duration of participation, because "need driven participants" left as soon as their need is satisfied.

In addition, existing literature has extensively discussed the role of experience in shaping individual participation in open source (e.g., Alexy & Reitzig, 2013; Bagozzi & Dholakia, 2006; Roberts et al., 2006; Von Krogh et al., 2003). Existing literature has argued that experience impact on contributor behavior, as a result of its imprinting effect on both intrinsic and extrinsic motivation. Alexy Alexy, Henkel, and Wallin (2013b) for example, studying a large multinational engineering firm, argued that previous experience with open source increases individual tendency to support open source, because of their familiarity with the community norms and ability to adopt open source with a lower learning cost in their own work. Bagozzi, & Dholakia (2006), in arguing the role of intrinsic motivation as discussed above, also found that such positive effect of social identification is reinforced by contributors' experience in the Linux user group, because their involvement and interaction with other members deepen as they gain experience in the community.

2.2. Extrinsic motivation

Research on open source also extensively discussed the role of extrinsic motivation of contributors in open source (e.g., Alexy et al., 2013b; Krishnamurthy et al., 2014; West, 2003). Compared with the literature on contributors' intrinsic motivation that focuses on the psychological attributes, such as emotion, cognition, self-identify and belief, the extrinsic motivation perspective emphasizes the reward system in open source and how the individual difference in the utility of such reward manifest the heterogeneity in contributing behavior. Reward system in open source seems to be paradoxical, because it strongly emphasizes the public goods nature of knowledge and encourages voluntary sharing (Hippel & Krogh, 2003). However, scholars have uncovered several non-pecuniary reward mechanisms that can also account for the contributing behavior in open source. The first external motivation is developers' career concerns. For example, Lerner and Tirole (2005a) explored contributor's motivation and behavior through the cases of Apache, Perl, and Sendmail. They summarized that programmers are motivated by the "the career concern incentives" about "future job offers, shares in commercial open source-based companies, or future access to the venture capital market" (Lerner and Triole, 2005a:14), because such participation in open source allows the performance and efforts of developers to be more visible to "relevant audience (peers, labor market, venture capital community)". In other words, working on open innovation technologies serves as a bridge for future financial rewards, and access to potential tacit knowledge to claim future intellectual property rights? Similarly, in a study of a large-scale telecom company, Alexy et al. (2013b) also found that the change of job role for programs, in favor or against open source innovation, would change the individual support of open source, because, for developers, such support can become the cue that allows them to fit in their organizational environment.

Research also investigated the design of monetary rewards in simulating contributing behavior, which also falls into the category of extrinsic motivation. Studying survey and archival data of the Apache project, Roberts et al. (2006) found that being paid can increase developers' participation in open source, as such extrinsic motivation can promote their perceived importance of the tasks. Krishnamurthy et al. (2014), noticed the particular importance of ideology in open source, found that developers are less motivated by monetary rewards in they hold a strong belief in the open source movement. In this case, they regard monetary rewards as a contamination of their belief and purpose.

Lastly, literature also discusses social norms and expectation of reciprocity that extrinsic motivate contributors. For example, open source was also considered by scholars as a type of "gift economy" (Zeitlyn, 2003). Theorizing from a sociology and anthropology perspective, Zeitlyn (2003) argues that contributing behavior is essentially a gift-giving activity with the expectation of reciprocity, in the process of which contributors can accumulate "symbolic capital" that can cash out later for their own technological needs or career advancement. From a slightly different perspective, several studies investigate how open source contribution is motivated by the status seeking incentives for individuals that try to comply to the social norm of open source. For example, Roberts, Hann & Slaughter (2006) found that contributors' status seeking intention positively influence contribution in the Apache project. They argue that the need for status motive developers to showcase their talents through contribution, to gain recognition to the relevant audience.

2.3. Technological characteristics

Lastly, characteristics of open source technologies also shape contributor motivation and their resulting behavior (e.g., Belenzon & Schankerman, 2015; Foss et al., 2016; Oh & Jeon, 2007; Shah, 2006). The first technology level characteristics is the size of open source communities. Larger open source projects encourage contributors with extrinsic motivation, because it allows higher visibility and more effective status-enhancing (Belenzon & Schankerman, 2015). The network externality due to size can also benefit the learning process of contributors (Oh & Jeon, 2007). Another technology-related factor is the openness of the project (e.g., Belenzon & Schankerman, 2015; Shah, 2006). Although open source entails full disclosure and open communication, innovation in such mode can still vary in openness in terms of their control over the innovation process and the outcome of development. For example, studying the mail list of two major open source community, Shah (2006) found that owner’s control of decision rights decreases external contribution, because it lowers potential contributors’ expectation that the open source technology can meet the heterogeneous needs of users and flexibility in the content of contribution. In a more recent study based on 149,956 unique developers from souceforge.com, Belenzon & Schankerman (2015) further argue that the openness of projects impact on the type of contributor that can attract. Measuring openness through open source license as the extent to which the project forbids derivative commercialization (greater openness), they found that project with high openness attracts more anonymous and open developers, because of the alignment between user motivation with the ideology of the project. Meanwhile, existing research shows that the characteristics of the innovation process in open source also interacts with the contributor behavior. Also studying data from sourceforge.net, Foss et al. (2016) found that contributors are more likely to join projects with artifact-based communication that gears toward specific problem-solving (such as patch files, bug reports, etc) compared with those that rely on open-ended discussion, because artifact-based communication eases the barrier of contribution by explicating specifying subproblems that contributor can solve.

The innovation processes in open source

The second focus in open source research is the development processes. Compared with the motivation that highlights individual-level drivers underlying open source communities, this stream of research focuses more on the community level dynamics. The central question of interest is how innovation is created in open source and how to best manage the open source communities (e.g., Baldwin & Clark, 2006; O'Mahony & Ferraro, 2007; Stam, 2009; Von Krogh & Von Hippel, 2006). Because open source entails knowledge creation to transcend beyond firm boundaries, which is at the same time, no longer protected with prosperity rights (Haefliger, Von Krogh, & Spaeth, 2008; Lakhani & Von Hippel, 2003), scholars have long suspected that the processes and dynamics given rise to open source technologies would differ from that depicted in the innovation research based on commercial technologies.

3.1. Coordination

The first unique process discussed in the literature is the coordination dynamics with contributors (e.g., Dahlander & Magnusson, 2005; Lee & Cole, 2003; O'Mahony & Ferraro, 2007). Because open source innovation relies heavily on the commitment and knowledge of volunteering contributors, how to coordinate, manage and motivate contributors for sustained contribution becomes a critical and unique issue for open source technologies. In practice, coordination in open source is achieved through a community-based model (Lee & Cole, 2003). Lee and Cole (2003) summarized that the community-based innovation process fundamentally differs from the firm-based innovation model, because it does not restrict membership within organizational boundary and knowledge creation and retention can be extremely distributed. Studying the Linux development community, they found that the growth of the community in distributed innovation is essentially organized by an evolutionary process of learning from errors and imperfections. In addition, studies explored the management of open source communities. Dahlander and Magnusson (2005) discussed how companies with open source project should balance between their commercial purpose and the communities. Through a comparative case study of four European firms with open source projects, they found that manager invoke several strategies to coordinate with the crowd in open source for more effective innovation, including participating the community discussion, maintain reputation, monetary rewards for problem-solving, creating online-forums and mail lists etc. Those coordination strategies then were theorized into three distinct categories of strategies, named as “symbiotic”, “commensalistic”, and “parasitic” approaches to handle open source communities as proposed by the authors.

3.2. Communication

Accompanied with coordination, communication constitutes another important focus in studying the innovation process in open source. While traditional firm-based innovation relies mostly on daily face to face interaction and communication (Cohen & Levinthal, 1990; Nelson & Winter, 1982), open source are dominated by technology-mediated communication, through online forums, email lists, etc (Bagozzi & Dholakia, 2006; Dahlander & Frederiksen, 2012; Lee & Cole, 2003). Addressing such distinct communication, studies have explored how different types of communication could impact on community dynamics and contributors’ behavior. Studying open source projects from sourceforge.net, Foss et al. (2016) investigated how the two types of communication emerged in open source, open-ended communication and artifact-based communication based on problems, lead to initiating and contributing behavior in the community. They argue that open-ended communication is essentially a process of problem-formation, and hence can lead to the creation of new projects. On the other hand, artifact-based communication manifests problem-solving and hence will attract more contributing behavior. In addition, frequent communication is also found to give rise to the emergence of lead contributors, because the heterogeneity of open source contributors further reinforces the importance of communication in stimulating other’s motivation and keep interests aligned (O'Mahony & Ferraro, 2007). Similarly, strategic interaction with can also structures of interaction through reciprocal activities by establishing linkages among contributors (Kuk, 2006).

3.3. Control

A third factor discussed in the existing literature is the control over open source communities (e.g., Alexy, George, & Salter, 2013a; Henkel, 2006; Kogut & Metiu, 2001; Kuk, 2006). The first is owners’ control over knowledge sharing. Although open source entails transparency of source codes, it does not mean that firms are required to disclose an entire innovation in open source. To some degree, the knowledge reveals of the firm as a control over community is a firm-level equivalent to contributor behavior. Firms control the decision rights of disclosure and open source content. Such possibility brings the question, what drives the extent to which firms will disclose through open source and what are the consequences? In a theory paper, Alexy, George & Salter (2013) discussed how technological uncertainty, knowledge structure and the value capture potential of a technology affect the firm’s decision of revealing knowledge. In the context of open source, some of those have been supported by empirical evidence. For example, studying the case of Linux, Henkel (2006) found that the code sharing activity is positively related to the need for technological supported and proprietary complementary assets, because, in such situations, open source lowers development costs without compromising the value capture opportunities backed by private complementary resources. In addition, inter-dependency of knowledge can also increase the tendency of knowledge reveal tendency. Based on the open source project of K Desktop Environment, Kuk (2006) reported knowledge sharing can be stimulated by the cross-thread connectivity in the mail list of the project, as interdependency allows higher chances that the crowd can extract useful information for improvement after revealing the knowledge. Existing literature also noticed the potential risks of excessive knowledge reveal. Kogut and Metiu (2001), for example, caution against the potential risks of "forking" in open source, as such replication of knowledge may create competing versions of technologies that erode the dominance of the originals.

Another consideration is the control over the decision rights during innovation. In study motivation of contributors, Shah (2006) noticed that the two sample communities vary in terms of who can overwrite the existing code and whether they allow different opinions to be voiced through the mail list. Such differences, as she then observed, lead to the variance in the level of contributions from hobbyists, in a way that tighter control reduces hobbyists' contribution.

The third control strategy is the institution of open source license. Open source license was initially created to specify ownership of open source innovation. Although by design, open source forfeits the proprietary rights of innovation, it does not mean that such innovation leaves ownership or intellectual property rights undefined (Comino, Manenti, & Parisi, 2007; Lerner & Tirole, 2005b). Open source licenses are essentially loosely enforced contracts attached with the disclosure of technology. Through open source licenses, the owner of a technology declares the authorization to the public regarding the use, distribution and modification of the technology, while retaining copyrights along with other requirements that vary across different forms of licenses. The earliest popular license was GNU General Public License created by Richard Stallman of the Free Software Foundation, one of the most famous pioneers of the open source movement. The GPL license enforces mandated disclosure of the subsequent innovation. That is, once an innovation incorporates open source technologies with the GPL license, it also has to be open sourced and with the same GPL license. In recent years, however, more liberal license such as BSD (Berkeley Software Distribution) and MIT license, which allows commercial derivatives without mandated disclosure, have become more dominant (Fitzgerald, 2006). Existing studies found that open source license constitutes an important control tool to protect open source innovation against private appropriation (O'Mahony, 2003). For example, qualitative studying 5 open source projects, O'Mahony (2003) found that open source license is frequently used to enforce legal and normative sanction on members while deviate from open source community norms, so as to deter the appropriation of open source technology for commercial users. In turn, they argued that such license reduces the involvement of commercial actors. Similarly, Henkel (2006) found that such institutionalize control over the knowledge disclosure can facilitate knowledge sharing from members in the context of Linux, because it mitigates the concern that knowledge shared in open source may be used for commercialization. Belenzon and Schankerman (2015) use open source license to measure the openness of the project and compare licenses with mandated disclosure (as high openness, such as GPL) with those that do not require so (closed, such as MIT). Consistent with prior research, they found that GPL licenses anonymous and open developers, because of the alignment of the ideology. Through formal modeling, Lerner and Tirole (2005b) also found that when the level of trust is high, the owner of the technology is more likely to adopt a permissive license that allows ex-post value appropriation (BSD or MIT), rather than GPL.

3.4. System design

Lastly, scholars also investigated how to best design the innovation structure in communities. For example, through formal modeling, Baldwin and Clark (2006) proved that a modular architecture can improve the efficiency of open source innovation while attenuating the potential risks of free riding, because such design allows independent search for optimal solutions within each module and hence increase the overall efficiency of the open source system. In a later study of Mozilla, the authors found that as the technology becomes increasingly open, the sponsor of the open source community design the technology in a way that is more modular to improve the efficiency of open source collaboration (Alexy et al., 2013a; MacCormack, Rusnak, & Baldwin, 2006). Studies have also investigated the effect of community structure on the performance of innovation. In general, excessive concentration of active contributors in all participates impede the performance of open source technology by reducing the motivation for others to share and contribute knowledge (e.g., Kuk, 2006).

Competitive dynamics

The literature on competitive dynamics highlights the rivalry between proprietary innovation and open source. To some extent, this stream of literature is most relevant to the traditional strategy and innovation research, because it regards open source as a competing or alternative knowledge sourcing mode for innovation (Waguespack & Fleming, 2009). With the external developers who can scrutinize the technology and correct errors, open source allows easier and more timely improvements in the development processes.

Studies on competitive dynamics are interested in the firm's decision between open source and closed innovation. That is, when should a firm choose open source? How should firms respond to the open source movement? And how can priced proprietary technologies compete with open source innovation that is distributed for free? In contrast with the innovation process research focuses exclusively on within open source communities, this literature is more interested in the comparison of different innovation models, as well as their technological and commercial consequences. This literature investigates both the antecedents and consequences of open source from the perspective of the firm, regarding closed firm-based innovation as alternative to each other.

1.4.1. Antecedence/decision of open source

Research on the antecedence of open source explored when open source can be more effective than other knowledge sourcing modes. Through simulation, Afuah and Tucci (2012) showed that crowdsourcing for innovation is a distinct governance mode in addition to internal sourcing, alliances or acquisitions. It should be the most effective for modularized problems with a distant solution that can be easily articulated, because the heterogeneity of the crowd increases the likelihood of obtaining optimum solutions outside the firm's knowledge domain. They specifically point out that open source is a subset of crowdsourcing (Afuah & Tucci, 2013). From a similar perspective, Almirall and Casadesus-Masanell (2010) theorized the impact of knowledge complexity and flexibility of changing partners on the choice of open versus closed innovation. Through simulation, they found that when there are a large number of flexible partners to solve complex problems, open innovation outperforms closed innovation, because it allows recombination of a large number of best solutions for each sub-problems in order to identify the optimum solution for the overall innovation. Felin and Zenger (2014) also argued that open innovation is superior in solving complex problems, because the extensive knowledge sharing can facilitate problem-solving by forming theories and heuristics to guide the distant search of a solution. In summary, those studies found that open source is particularly favorable when the technology requires distant and complicated knowledge.

The decision of open source can also be driven by the demand side factors. Henkel, Schöberl, and Alexy (2014), investigated how consumers can be an important driver of open source. Studying embedded component manufacturers based on Linux, they found that the choice of open source by companies can be motivated by customer demands. Disclosing source code stimulates demands, not only because it allows higher customizability and the ability to fix bugs, but also customers, who are often part of the Linux communities, are ahead of the firm's own adaption into the new technology of Linux. Hence, the open source also aids firms' adaptation and learning by meeting consumers' demands. Wen et al. (2015) investigated the impact of institutional uncertainty on open source innovation. Studying IBM's creation of patent commons and waiving litigation against open source communities in the 2000s, they found that lower litigation risks encourage new ventures' entry to open source technologies.

1.4.2. Performance of open vs. closed innovation

Literature also juxtaposed the innovation performance of open source with closed knowledge sourcing modes in studying innovation performance. To date, the limited results demonstrate that the performance of open source is highly sensitive to the competition proprietary innovation. Studying the impact of potential threats of IPR litigation, Wen, Forman, and Graham (2013) found that open source projects facing such risks are less likely to be adopted by users based on over 24,301 open source projects in Sourceforge.net. Such decrease of adoption is largely due to the potential of litigation induces increases the perceived cost of adoption, particularly when knowledge such infringement potential can be reused over time during the cumulative innovation in open source. In a simulation analysis, Bonaccorsi and Rossi (2003) considered a market with both commercial and open source software. Without the entrance of incumbents, open source software will become dominant in the market, unless commercial technologies can involve extensive R&D to compete on quality. However, under the present of incumbents, commercial technology can still take considerable market (38%). Through the similar formal modeling and simulation methodology, Economides and Katsamakas (2006) showed because vertically integrated property technology will outperform open source technology in terms of both market share and profitability. The synergy among integrated products created sticky demand that can substitute the need for open source technology that is often modularized. Such situation, however, will change, if the maximum potential demand for open source innovation is larger than the vertically integrated priced products.

Apart from competition, factors that preventing the high performance of open source innovation also emerge from the limited capability of organizations internally. Piezunka and Dahlander (2015) found that crowdsourcing may not achieve the intended technological benefits, because organizations are with limited attention span, and they tend to simplify and rationalize the filtering process based on their existing capabilities and knowledge. The analysis in suggestion forums for large incumbent manufacturing companies showed that even when firms can attract distant solutions, the often few to recognize their potential.

1.4.3. Value appropriation

The next important question in the competitive dynamics is related to value appropriation and capture in open source. It is of central interest for research on strategic management o address the question, even if open source is can produce more effective innovation and superior technological performance, how can firms without propriety rights? The first answer given in the current literature is selective revealing (e.g., Henkel, 2006). Firms only partially open source their knowledge to learn and develop absorptive capacity from the crowd, while keeping other knowledge that is more critical to value capture private. As evidenced in the studies summarized in the previous session, open source is indeed more likely under technological uncertainty, when the promise of value capture from a technology is ambiguous (Alexy et al., 2013a).

The second answer is the integration between open source and private technologies. Lerner and Tirole (2005a) first propose that there are several ways for companies to exploit open source, including providing priced complementary services and products, proactively waving proprietary rights, initiating open source platforms, etc. Similarly, Von Hippel and Krogh (2003) argued that open source innovation, in nature, is not entirely a collective action of a social movement as depicted in early research. Rather, it resembles more with a "private-collective mode" of innovation that contains both private investment and collective creation of knowledge as public goods. The private investment, often neglected by the literature, happens when the inventor was seeding the innovation before open source, and when the inventor/initiator of the technology offers monetary rewards or other incentives to the crowd for problem-solving. The collective creation knowledge, on the other hand, refers to the further refinement and development of technology after disclosure within the open source community. Alexy and Reitzig (2013) further investigated why firms are motivated to invest in technology for such "private-collection mode", even though it can severely undermine the value capture of technologies. They propose that, by purchasing exclusion rights on potential future innovations based on the open source technologies, innovators can reshape the appropriability regime of open source innovation. Using an exogenous shock of disclosure of potential patent infringement in open source, they found that firms that open sourced their innovation under the “private collection mode” is more likely to release the related patents as patent commons (pledge to waive exclusion rights) (e.g., IBM) than those rely on proprietary innovation (e.g., Microsoft) because of the mechanism reasoned above. Similarly, Fosfuri et al. (2008) found that publicly listed software companies with large stocks of intellectual property rights (e.g., patents) are more likely to launch open source-based products, as they have more power to control the innovation, more complementary assets and face lower threats of litigation. This finding echoes an early research (e.g., West, 2003), in which authors study how incumbents adopted a hybrid strategy of innovation in response to the emergence of Linux. West (2003), through multiple case studies in computer software and hardware incumbents, argued that the hybrid strategy can be achieved through establishing open standards and altering the terms of open source licensing that no longer restrict the ex-post value appropriation. In doing so, firms can retain control while reducing duplicative development efforts. Hence, such strategy is particularly preferable to firms that face competition from open source innovation. Bonaccorsi, Giannangeli, and Rossi (2006) also discussed such hybrid model through the survey of over 100 Italian software firms. Against the notion that firms can capture value by offering priced open source solution service, they found that firms rarely use it as a pure business model. Rather, depending on their experiences with open source, the business models of software firms are often a mixture of licensing revenue from proprietary innovation and service revenue from open source technologies.

Critique

In this section, I reviewed the existing studies on open source innovation. The literature to date still focuses on the three broad issues Von Krogh and Von Hippel (2003) raised in the special issue of open source innovation in 2003, namely (1) motivation of contributors (2) innovation processes (3) competitive dynamics. Table 1.1 provides a detailed summary of research open source innovation.

Insert Table 1.1 Here

Overall, most studies on open source innovation cast such phenomenon as a social movement rather than market competition. In turn, the studies on the motivation of contributors explore why individual developers are motivated to work on open source innovation without getting paid or financial rewards. To summarize, existing literature outlines three sets of factors underlying contributor's motivation to participate in open source communities. Different from inventors from a commercial setting, contributors are intrinsically motivated to participate because open source can reinforce their ideology of ant-commercialization knowledge, as well as their use of the technologies (Baldwin & Clark, 2006), as well as their own needs of the technology (e.g., Lakhani & Von Hippel, 2003; Von Hippel, 1986; Von Krogh et al., 2003). More specifically, frequent and lead contributors in open source often hold a strong belief in the ideology that knowledge should be public goods accessible to everyone, rather than sources of business profit (Levine & Prietula, 2013). Their innovation activities are also motivated by their own needs of using the technology (e.g., Jeppesen & Frederiksen, 2006). In later works, studies also noticed that some of the motivation underlying the labor force of such open innovation technologies share similarities with human capital in for-profit organizations. Despite the unique ideologies, contributors are still sensitive to deferred financial rewards, like boosting one's status and by extension one's career prospects (Baldwin & Clark, 2006; Shah, 2006), as the transparency of open source knowledge allows those activities to a strong signal to contributors' capabilities (Hertel et al., 2003; Roberts et al., 2006; Shah, 2006). Lastly, the contributors' experience and the characteristics of project can also play a role and interact with contributors' intrinsic and extrinsic motivation, in such that the more experience, project size, and openness all facilitate more active contribution (Alexy et al., 2013b; Foss et al., 2016; Von Krogh et al., 2003).

For the process of open innovation, studies investigated the unique innovation processes that govern the use and allocation of knowledge and human capital resources within individual open innovation, in terms of coordination, communication, control and organization/community structure (e.g., Baldwin & Clark, 2006; Boudreau, 2010; Dahlander & Frederiksen, 2012; O'Mahony & Ferraro, 2007). Studies reveal the open source innovation rely on a unique mode of development from several perspectives. Frist, the coordination and communication is basically technology-mediated, through online communities, mail list and forums (e.g., Lee & Cole, 2003; Von Krogh et al., 2003), rather than face-to-face interactions. Second, owners/initiators enforce control over open source communities mainly through selective reveal (Henkel, 2006; Henkel et al., 2014) and open source license (Lerner & Tirole, 2005b; Rosen, 2004). Such control over the outcome of innovation is enforced by disclosing information and unique copyright contracts that retain ownership but forfeits propriety (e.g., Fitzgerald, 2006; Lerner & Tirole, 2005b; Rosen, 2004) is distinct from the strategies and institutions implemented in commercial innovation. Lastly, the structure of the community can be fluid and constantly evolve, which exerts significant impacts on the effectiveness of open source innovation. Moreover, compared with the literature on the motivation of contributors, the discussion of the innovation process is at the technology level, and focuses on the performance implications of different dynamics and strategies in open source. While I seek to review the papers related to innovation by categorizing them into different pockets, the literature on open source innovation process is actually scattered. Each paper focuses on different elements. Although the four factors discussed should be tightly connected with each other, few studies investigate the interaction and interdependence among those processes. Hence, a systematic process model that depicts on open source innovation is missing in the current literature. Also, at the level, the unit of analysis is technology and community, without differentiating firms/ventures who initiate such innovation from the user/developer community that rely on and maintain the development of the technologies.

Lastly, the competitive dynamics literature emphasizes the performance implication of open innovation. Considerable research focuses on the co-existence of private and open source during innovation and competition, with the emphasis on the role of litigation risks (e.g., Wen et al., 2013), complementary assets (e.g., Fosfuri et al., 2008), and the process of technology diffusion (e.g., Boudreau & Jeppesen, 2015). . At the same time, studies have also focused on the choice of open source over proprietary innovation explored the role of knowledge structure and characteristics (e.g., complexity, modularity) (e.g., Almirall & Casadesus-Masanell, 2010; Boudreau & Jeppesen, 2014), highlighting the tradeoff between distance search and extended reveal of knowledge. In those investigations, open source is either regarded as an alternative knowledge sourcing mode, or potential competitors of commercial closed innovation. Studies have found that open source is more likely for modularized technologies that are distant to the organization (Afuah & Tucci, 2012), and when potential users demand the high flexibility of modification (Lakhani & Von Hippel, 2003). Other studies, in contrast, emphasize the possibility of integration between open source and proprietary innovation, in which open source is regarded as a "private-collective" innovation. Private investment can happen before the technology is open sourced, making open source as a "private-collective" mode of innovation, rather than an anti-commercialization social movement. In doing so, firms can resort to the crowd for distant search while capture value from complementary assets, services and other business models (Alexy & Reitzig, 2013; Hippel & Krogh, 2003). While most studies on the contributors and innovation process within communities highlight the non-profit and public-goods nature of open source innovation and portrait open source as a social movement, the literature in this stream does not assume the independence of innovation with the business world. Rather, it argues that open source can be an alternative sourcing mode for knowledge creation, in a way that is similar to other outsourcing modes (e.g., Afuah & Tucci, 2012; Howe, 2008; Piezunka & Dahlander, 2015). To a large extent, this literature is most relevant to business and strategy research, as it focuses on competition and value capture. The rich discussion about the antecedence of open source as knowledge sourcing mode and the integration of open source with private innovation provided detailed answers to when and why for-profit companies are willing to open source and give-up proprietary rights.

Together, these three pockets of research on open source constitute a very detailed delineation of open source innovation. In summary, open source innovation is fundamentally distinct because its unique underlying ideology that highlights the "freedom" or public good nature of knowledge. In essence, open source innovation is highly cumulative and problem-solving oriented (Felin & Zenger, 2014; Foss et al., 2016). Open source innovation process is decentralized and fluid, characterized with extensive involvement with external actors, in an environment where development activities are regulated based on unique coordination, communication and control (Almirall & Casadesus-Masanell, 2010; Lee & Cole, 2003; Lerner & Tirole). Meanwhile, knowledge is distributed largely outside firm boundaries. The emergence of open source has also affected commercial innovation, forcing firms, especially incumbents to respond through competition and integration. At the same time, the questions unaddressed in the existing literature provide ample research opportunities for future studies.

Research opportunity 1: the problem of inventor retention in open innovation. The unique novation and diversity of contributions in open innovation give rise to the problem of retention, a unique challenge in managing open source community. One the one hand, as those contributors are driven by the need to better use the technology are likely to leave when their demands are satisfied by the improvement (Bagozzi & Dholakia, 2006; Baldwin & von Hippel, 2011), containing the fluidity of participation becomes critical to the sustainability of open source innovation. On the other hand, other types of contributors that are driven by intrinsic motivations and the belief that open innovation is a social movement are less likely to respond to the traditional incentive structure to retain human capital resources in a traditional organizational setting. Accordingly, how to incentivize them to stay while cultivating committed hobbyist contributors (Shah, 2006) becomes a challenge to open source-based organizations and communities. The dilemma of maintaining contributor's commitment is further aggravated by the fact that most effective contribution originates from the problem-solving process that attracts need/demand-based contributors that have a tendency to leave after finishing the problem-solving. One way is to have sustained contributions through different individuals over time, another is to have sustained contributions within-contributors over time. Such choice of gaining sustained collaboration can vary systematically across platforms, and across entities behind the problems worked on through open source. Yet, the current literature has not investigated the implications of each possible strategies and provided a satisfying answer to such tension theoretically nor empirically.

Research opportunity 2: the heterogeneity of collaboration across open innovation and its impact on innovation. Related to the previous point, existing studies fall short in explaining the heterogeneity of contribution across different open source projects. That is, why do contributors join certain open source projects over others? While emerging literature has explored some project-level characteristics, it has yet explored detailed mechanisms underlying the effect of such features in attracting contribution. Moreover, current literature has not addressed whether such heterogeneity is a result of selection of external contributors or can be a strategic choice of the initiator of open innovation. Indeed, contrary to the depiction of open innovation in the current literature, collaboration may not be always desired, due to its potential risks related to the labor retention and path-dependencies. Such inquiry to whether initiator may seek to contain collaboration and its possible consequences is particularly relevant, given the drastic heterogeneity in terms of the number of external contributor projects can attract in open source. Moreover, the consequences of such external collaboration on innovation performance are underexplored. As existing research largely focuses on the motivation as antecedents of contribution, it remains unclear whether contributors can indeed positively influence the open source innovation, technologically or financially. In fact, the positive effect of contributors is almost assumed in all the studies in this vein. Yet, there are reasons to suspect some dark side of contributors. For example, the fluid participation may disrupt the routines of innovation. The heterogeneous demands by contributors of the functionality of innovation can also generate the risk of hijacking the direction of the innovation from the initiator. Current research has not addressed such possibilities and provide clues on how open source-based organizations can reconcile such potential conflicts between the stability of innovation process and heterogeneity of contributor motivation.

Research opportunity 3: the evolution of the institutional environment in open innovation is not addressed. More specifically, this under-addressed issue is related to the open source license, the loosely defined contracts between the initiator of open innovation and the users and collaborators. Open source license is conceptually relevant particularly to venture boundary decisions because it resembles contractual agreements in a traditional market environment, which could substantially affect the anticipated transaction cost of collaboration and utilization of open and public knowledge. Although long noticed that open source license is a critical institution that facilitates the rises of open source innovation, most of the research is largely based on the GPL license, the most popular license early on, especially due to the adoption of GPL by Linux. Yet, in recent years, the forms of open source license have become increasingly diverse. Apart from authorization to the public for free usage, they differ substantially in terms of the retention of trademark, copyright, disclosure and commercialization for subsequent derivative work. Moreover, the dominance of GPL license has been gradually eroded by permissive and closed-ended licenses such as MIT, which allows the user to commercialize their own innovation based on the focal technologies. It not only manifests an institutional change worth studying the context of technology and innovation but also allows the possibility of value capture from open source innovation, which is forbidden by the GPL license.

Relatedly, existing literature has not explored in detail how open source license is enforced and the possible consequences (Lerner & Tirole, 2005b). To date, there is considerable legal ambiguity about the nature of open source license as an enforceable contract (Rosen, 2004). Yet, if the likelihood of enforcement is low, then how can open source license allow the sponsor to maintain control over the innovation processes? Future research may explore the impact and mechanisms of open source license following those directions, so as to develop a better understanding about how the institutionalization of knowledge disclosure can shape the process and outcome of open source technologies.

Research opportunity 4: the existing research has yet investigated the role of communication technology in the coordination and communication of open source. Although the technology-mediated nature of communication is noted by the literature, few studies investigated the evolution of communication technology over time in communities. In practice, the communication and coordination medium for open source innovation has shifted from the discussion forum or mailing lists (e.g., Raymond, 2001; Shah, 2006), to various version control tools (e.g., Belenzon & Schankerman, 2015; Wen et al., 2013), largely dedicatedly developed for the purpose of coordination in open source. Compare with discussion forums or mail lists, participants need to verbalize their ideas or problems they encounter, version control allows contributors to directly coordinate on the technology. It allows contributors to work on and compare different branches of open source technologies, while owners to make the decision of changes by external contributors. Indeed, the archival data used in the empirical analysis of several papers reviewed in this section are based on version control tools (sourceforge.net) (Foss et al., 2016; Wen et al., 2015), while another git-based version control tool, has given rise to GitHub, currently hosting the largest number of open source technologies (Dabbish, Stuart, Tsay, & Herbsleb, 2012). The current literature has not addressed how the transition from verbalization of ideas to direct coordination in source codes, enabled by the development of coordination technology, could impact on the innovation process in open source. It also remains to be explored whether sponsors would maintain the traditional mailing list-based coordination and communication after the emergence of new technologies, and what are the impacts of maintaining multiple channels for coordination and communication can impact the outcome of innovation and the dynamics within the communities. Those questions require further inquiries.

Research opportunity 5: the assumption of open source as a knowledge sourcing mode may require more detailed examination. In particular, considerable studies contend that the reveal of knowledge in open source to gain additional input to further develop the innovation. Yet, in reality, the absolute majority of open source projects fails to attract any participation (Octoverse, 2018). Such concern that open source for collaboration and knowledge sourcing may not be effective can be further aggravated as open source innovation does not entail ex-ante contract with the crowd in terms of their responsibility of knowledge creation. If firms are aware of the difficulties in knowledge creation through open source, then, would such consideration alter the current conclusion that is drawn based only on knowledge characteristics? One possibility is that apart from knowledge sourcing, open source is motivated by other needs of the firm. For example, studies using the framework of the "private-collective" mode theorize that firms seek to attract more contributors to maintain the development of innovation after private investment. The alternative, yet addressed in the existing literature, is that the sequence of private investment and open source, could be the other way around. Firms initiate and cultivate open source innovation, and they seek to identify the opportunities that worth investing as proprietary innovation. In doing so, it is possible that rather than knowledge sourcing for subsequent development, open source can become the antecedence of proprietary technology as firms utilize open source as an experimentation to explore technologies and market before deciding on the proprietary investment. Those possibilities require further investigation in subsequent studies.

Research opportunity 6, the technology and financial consequences of open source remain largely under-explored. Most the literature on open source focuses on the antecedence related questions, such as the individual motivation to participate, and firms' choice between open source and private external knowledge sourcing mode. The prospect of open source in product market is rarely investigated. We still lack understanding on how open source can influence the technological performance. Would open source lead to technology superior innovation? How open source alter the diffusion of technologies compared with closed innovation? When technologies are provided for free, without price mechanisms, how do users choose among all potential alternatives? All those questions require more comprehensive studies in future research. Relatedly, the competition among open source technologies and within community dynamics is not integrated. In studying the competition between proprietary and open source innovation, the implicit focus is the mode innovation, rather than specific technologies. The few studies that focus on the emergence of dominant technologies provide a confusing picture that runs against the reality. For example, the simulation studies of Bonaccorsi and Rossi (2003) and Economides and Katsamakas (2006) predicted the proprietary products by incumbents can still take considerable market share, even outcompete open source, because of the synergy and non-substitutability such products can create. It seems to be a reasonable conclusion given the dominance of proprietary software like Windows and Microsoft office. However, such conclusion runs counter to the fact that Linux can still dominant in presence of Windows in the server operating systems for supercomputers (Dua, 2017).

The conflict between the existing findings and reality gives rise to at least two set of questions for future research. From the perspective of open source, how open source innovation can overcome the lack of synergy, compared with proprietary integrated products, in technology competition? Is it because open source enables high technological performance through distance search, or it is related to its advantage in diffusion due to its free nature? The second question is how incumbents respond to the threat from open source. From the above anecdotal evidence, it seems that open source does have a technological advantage, as supercomputer requires the highest technological performance of the operating system. Then, how do incumbents adapt when technological change and evolution has become increasingly rapid and transparent through open source innovation? Can incumbents keep pace and learn from open source through proprietary patent-based innovation? Relatedly, what are the consequences to themselves and their incumbent competitors, if they decide to open source their technologies under the pressure of such trends? Interestingly, it seems that the open source projects by incumbents have attracted considerable interest in open source communities (Octoverse, 2018). However, to date, no studies explored whether and how open source allows incumbents to better adapt or maintain advantages in competition.

Research opportunity 7, how open source technologies compete with each other. Linux was initially just one of many versions of open source Unix (Techworm, 2016). Although the literature frequently investigated Linux communities, it rarely explored the question, why Linux took the dominance, among other alternatives. In other words, existing literature has not investigated what strategies open source innovation can leverage in the competition with other open source technologies. Relatedly, we lack the understanding of how the rise of the dominant design and the overall technology evolution is altered by open source, in competition with proprietary technologies and among themselves. Yet, there are reasons to suspect the open source nature would somehow impact on the current conclusion of technological competition, which is based on the assumption that firms strive to protect the knowledge that can give them competitive advantages.

Research opportunity 8, existing literature rarely discussed entrepreneurship and new ventures based on open source technologies. Currently, the discussion on the competitive dynamics focused on the tension between open source and incumbents such as IBM and Microsoft. How open source can impact on entrepreneurship, which rapidly populates open source, in contrast, is rarely discussed. Compared with incumbents, new ventures are more likely to be both the users and innovators in open source as they face severe resource constraints and other liability of newness. They are more likely to use open source technologies to run their business for cost saving purposes. Meanwhile, they are also more likely to resort to open source for knowledge and experimentation as they lack the capabilities and resource to invest in proprietary technologies. The challenge, yet addressed in the existing studies, is whether and how open source innovation allows new ventures to gain financial benefit and revenue to survive and grow? All those questions require further investigation.

Meanwhile, the link between corporate strategy and open innovation in general need to be further strengthened. As open source was initially conceptualized as the independent content of commercial business based on technology and innovation, considerable early research focuses on the technological impact and development of communities in a way that does not have substantial implications to corporate strategies. However, as already noticed by a few studies, open source has been increasingly integrated into the business world, with increasingly more firms become an active participant in open source. Existing studies have demonstrated that incumbents could participate open source to ease competition and facilitate adaptation (e.g., Waguespack & Fleming, 2009), while open source also provides resource-constraint new ventures knowledge input and the market for experimentation at very low cost. Yet, despite such possibilities, existing literature rarely addresses how open source is strategically used by incumbents and new ventures. It is not known how participating in open source shape the competitive dynamics among different types of firms in a same technological field. Moreover, we still don’t know whether open source a mode of innovation can shape the technology evolution as they alter the basic mechanism of knowledge creation and protection in the competition of dominant design, while the cumulative nature of innovation can also impact on the emergence of radical technologies. In terms of theory development, studies on open source focus largely on the phenomenon, without invoking much of the theories that are fundamental to strategy and innovation in commercial settings. However, the need for research on open source, drawn and adapted from established management theories is particularly relevant and needed, given that the landscape of open source has substantially changed over the decade. While it is increasingly populated by new ventures, incumbent firms like Microsoft, once the open source movement was against, has taken the lead in open source (Octoverse, 2018). The corporate participants that operate largely in accordance with the established management theory also provide an opportunity to bridge open source as a newly emerged phenomenon with the theoretical development in management, strategy, and innovation.

Relatedly, existing literature on open source innovation has not discussed open source innovation impact on the evolution of technologies fields and technology cycles, a central focus in the discussion of the traditional firm-based innovation (e.g., Adner & Kapoor, 2016a; Anderson & Tushman, 1990; Dosi, 1982). Are the technologies developed in open source radical in nature, or incremental improvement of established technologies? Or nature of the innovation can vary across different owners? We know little on how the innovation process in open source technologies may shape nature of the technologies in the process of development and how each element reviewed above plays a role. Most fundamentally, would the full disclosure of knowledge beyond firm boundary affect the evolution of technology in new ventures and incumbents differently? The research on open source innovation process has yet addressed issues related to those questions and explored how the nature of the innovation interacts with the unique innovation process through open source.

Research opportunity 8, quantitative studies based on large data with a methodology that can derive causality is needed. It should be noted some limitations of current research are also associated with the methodology used in the current literature. The majority of the studies reviewed here is either qualitative or survey based on a small sample. Such methodology makes it more difficult to gauge the outcome related implications and to probe for underlying mechanisms driving the effect observed in the studies. As we have come to the era of big data and the open source activities become increasingly digitalized, large archival data on open source has become more available. Such new trends in data accessibility allow many questions related to heterogeneity across projects to be better explored. It also provides opportunities to better juxtapose proprietary and open source at technology level to develop a deeper understanding of the competitive dynamics of open source.

TECHNOLOGY PLATFORMS

Another literature that is relevant to this dissertation is the research on platform-based markets. In this stream of literature, platforms are essentially defined as "a product …when it is one component or subsystem of an evolving technological system, when it is strongly functionally interdependent with most of the other components of this system, and when end-user demand is for the overall system, so that there is no demand for components when they are isolated from the overall system." (Gawer & Cusumano, 2002: 2008). In this section, I briefly review the literature on platform-based innovation, with a particular focus on the cross-platform strategy. I first elucidate the idea of externality, a key definition and constructs in theories underlying the research on technology platforms and how open source can be regarded as platform-based technologies. Then I discuss briefly the central foci of this literature. Lastly, I focus on the cross-platform dynamics, a particularly relevant topic to the competition among complementary open source technologies through the review of multihoming on platforms.

Network externality

The idea of platform-based technology was initially brought up by Katz and Shapiro (e.g., 1986, 1994) when discussing the effect of network externality in technology adoption. They first noted that many industries and products strong network externality, in which "the benefit that a consumer derives from the use of a good often depends on the number of other consumers purchasing compatible items" (Katz & Shapiro, 1986: 823). Hence, the adoption of technologies by potential users hinges upon the extent to which they are connected to the existing users of the technology (Suarez, 2005). If potential users are tightly connected with a large number of existing users, they are more likely to join as they can derive more benefits and higher utilize by reinforcing such connections. Such effect is the direct network effect. A typical example often used in the literature to users' choice of carrier in the telecom industry (McIntyre & Srinivasan, 2017). The more a person's contacts chose the carrier, the more benefit the potential user can obtain by joining the same carrier. Another type of network effects coming from the supply side of the platforms, that is users can also benefit from the addition of more functionalities or services attached to the platforms, which is usually provided by third-parties and referred in the literature as indirect network effect (Farrell & Klemperer, 2007). A typical example of indirect network effect exists in operating systems, the more software developer supply different technologies, the more user can drive utility by using the operating system and the software they demand. Vice versa, software providers also benefit from more users adopting the operating system, because it increases the overall market of their software (Farrell & Klemperer, 2007).

One distinct feature, implied in the discussion of network effects, is the nature of two-sided market of platforms (Armstrong, 2006; Farrell & Klemperer, 2007). Indeed, the network effects make platforms a unique governance mode for transactions between the supply side and the demand side, whereas in the regular market, the utility of transactions is considered as independent.

More specifically, the two-sided nature of the platform composed of three critical elements, the platform infrastructure, buyers and suppliers (McIntyre & Srinivasan, 2017; Rochet & Tirole, 2003). The platform infrastructure, in a high technology setting, provides basic technology framework that specifies the knowledge creation routine and standardized procedures of supply. In essence, many open standards and technology committees (Ranganathan & Rosenkopf, 2014; Waguespack & Fleming, 2009), can be regarded as such as platform infrastructure. Most often, especially in proprietary settings platform owners have considerable power in setting up the infrastructure. Suppliers, based on the infrastructure, seek to gain financial benefits by supply "add-on" functionalities to the infrastructure in a way that improve the overall technology. They are sometimes referred to as complementors and those "add-on" technologies are referred to as "complements". Then, customers choose to adopt the platforms they seek to transact the satisfy their own utilities. Customers or users are sometimes called installed base from the perspective of platforms. In such process, platforms functions as a medium that allow complementor to standardize their technological and transact with buyers (Evans, 2003; Rochet & Tirole, 2006), gaining profit from direct sales and growth of the platforms. It should be noted, although initially focus on technologies and innovation, later literature has extended the setting beyond high technology industries. It is argued that shopping mores, e-commerce services (like e-bay) and sharing economy (like share rides and Airbnb hotels) are all platforms that fit such conceptualization (Armstrong & Wright, 2007). Consequently, most of the investigations center on the market dynamics, rather innovation.

Platform competition

With the emphasis of network externality of platform-based technologies, the inter-platform competition is one of the most important foci in this stream of literature (e.g., Carrillo &

Tan, 2006; Rochet & Tirole, 2003; Shapiro & Varian, 1998). Such focus is rooted in the question, how network effects (or externalities) impact on the market competition process on platforms. Katz and Shapiro (e.g., 1986, 1994) first noted that in situations with significant network effects, the sequence of entry is extremely important. As later user's adoption choice is affected by the overall users on competing platforms, they are more likely to choose the one that first gains large installed base first (Suarez 2005). Those who entered the market first will have substantial advantage and gain a positive feedback loop, which ultimately results in a winner take all (WTA) situation (Evans, 2003; Katz & Shapiro, 1994; Liebowitz & Margolis, 1994; MacCormack et al., 2006; Rangan & Adner, 2001). Such situation can be further aggravated when platform owners can enforce switching cost on users, the expenses incur to users when they transition from the current platform to a new platform (Hagiu, 2006). Typical examples include the price customers has to pay for video consoles or cell phone when switching video game platforms or telecom carriers/mobile operating systems (e.g., Cennamo & Santalo, 2013; Farrell & Klemperer, 2007; Katz & Shapiro, 1994).

The extent to which that platform-based competition can result in the WTA situation can also vary. For example, through simulation, Lee, Lee, and Lee (2006) demonstrated that WTA becomes less likely when install base become segmented and users only interact with part of the suppliers and users within the network. Such localized network effect implies demand and supply heterogeneity and hence weaken the effect of switch cost in preventing user migration across competing platforms. Similarly, Cennamo and Santalo (2010) found that in the video game industry, platforms with non-exclusive complementary technologies gain higher technological performance as they can avoid adverse selection in the exclusivity contract, providing the evidence that WTA may not be the universal outcome for platform competition. In terms of the consequences, in addition to the result of the monopolistic market structure, the WTA outcome also affects the technological adaptation of platform owners. For example, Schilling (2002b), studying several industries with network effects, found that later entrants displayed a higher tendency of technology lockout because of the lack of complementary technology and assets in the absence of indirect network effects.

Platform entry

Another related topic is the entry to the platform-based markets. A diverse literature has investigated entry in platforms from several perspectives. First, several studies explored who enter the platform-based market. For example, studying the video game industry, Zhu and Iansiti (2012) found that the entry success of video game consoles is affected by the relative importance of performance quality, as it shapes the importance of network externality and hence user's expectation of WTA when deciding which platform to adopt. Eisenmann, Parker, and Van Alstyne (2011) theorized how the entry to platform-based market can be motivated by the bundling of a complementary platform or weak substitute platform when the potential user based has high overlap with incumbents. They argue that such effects are due to the increased user net utility associated with such bundling that leads to the high possibility of switching to existing users.

The second stream of literature focuses on entry as suppliers or complementors. For example, Venkatraman and Lee (2004) found that complementors' entry to a specific platform is more frequent with low the network density, knowledge interdependency and emerging platforms in the U.S. video game industry. In those situations, complementors can easily differentiate themselves and gain competitive advantage. In a case study of Intel's platform strategy, Gawer and Henderson (2007) found that platform's owner's entry to the complementary market, providing complements directly, is affected by the belief in its ability to capture value on the complementary market. Lastly, a few studies investigated how platform's entry impacts on the overall development of the industry. Seamans and Zhu (2014) studied how the newspaper industry is affected by the entrance of platform-based craigslist and found that local newspapers suffered lower subscription rate due to the entry of craigslist that provide the publication of information for free.

Platform governance

The third focus is platform governance and control. The center of this research is whether platform should allow free complementor entrance, or grant access to evaluation (Boudreau, 2010). Boudreau (2010) argues that platform owners can either choose to grant access to complementors, or they can allow free entrance by giving up control over the access to the platform infrastructure. In the context of handheld computing systems from 1990 to 2004, he found that granting access stimulates the innovation rate of complementors five times as high as its alternative. He argued that such effect is large because granting access intensify competition event outside the platform, motivating complementors to innovate at a more rapid rate in order to get in. In contrast, Parker and Alstyne (2017) investigated the performance implications of openness of the platform. Through standard Cobb–Douglas production modeling, they show that open platforms without entry barriers and intellectual property protection can be profitable because it allows firms to better capture profit from ecosystem rather than direct sales. Hagiu, Wright, Andrei Hagiu, Hagiu Julian Wright, and Hagiu (2018) compared platform as a governance mode with vertical integration. Through interpreting formal modes in the context of professional services, they show that in such choice of governance generate a tradeoff between firms' need to generate "spillovers across professionals (best achieved by a vertical integrated firm)" (p.1) and their need to motivate professionals and "ensure professionals adapt their decisions to their private information" (p.1). In such situation, vertical integration is preferred when bonuses and variable fees is not feasible, platform-based governance is preferred because of their better ability to motivate the participants.

Multihoming on platforms

In studying platforms, the existing research noticed that complementors may not be exclusive to one platform, but rather provide technologies or services to multiple platforms competing which each other. An extended literature has discussed the implication of non-exclusivity of complementors in the platform competition. In some studies, such phenomenon is referred as complementor multihoming (e.g., Armstrong & Wright, 2007; Cennamo, Ozalp, & Kretschmer, 2018; Hagiu, 2009; Rasch, 2007).

Complementor multihoming, in which complementors participate in multiple platforms, however, could substantially change the outcome of platform competition. Caillaud and Jullien (2003) first noticed that, in two-sided markets mediated by platform-based information technologies (such as B2C platforms like Amzon.com and eBay), both complementors and users have the incentive to register with different intermediaries to contact more user base or expand their search of complementor service and information. They modeled the situation in which complementors are not restricted to one platform and can provide non-exclusive services and showed that the winter-take-tall structure may not emerge when multihoming is allowed. Rather, platforms can co-exist in an equilibrium state, if either complementors or users decide to multihome on several platforms. The implications of complementor multihoming to platforms is also discussed by Armstrong and Wright (2007) as "competitive bottlenecks". Authors argue that, when users are single-homing, multihoming complementors reduces the monopolistic power of platforms. While platforms have the incentive to charge higher prices to multihoming complementors for exclusive access of their users, such pricing strategy also drives complementors away and in turn deplete the indirect network externalities users can benefit. Hagiu (2006) further modeled situations in which entrance of complementors and users are sequential and proved that under the presence of complementor multihoming, it is possible for both competing platforms to make profits if platforms commit to an ex-ant price for users. It is also shown that when multihoming creates economies of scale across multiple, platform's price-cutting strategies will become less effective (Hagiu, 2009).

Empirically, Corts and Lederman (2008) provided insights into the increasing competition as a result of complementor multihoming. Based on data from video game industry, they found that multihoming of software provides generate cross-platform spillover of the indirect network effect, in such that platforms can also benefit from the growth of users in competing platforms. Also with data from the video game industry, Cennamo and Santalo (2013) also report that high overlapping of complementors in the same industry has a positive impact on market share of a focal video platform. In summary, those studies as provided strong theoretical evidence that complementor multihoming affects the outcome of a variety of strategies for platform owners and lead to increasingly heated platform competition.

Current research has also modeled the impact of complementor multihoming on users. A major conclusion from existing studies is that multihoming of one-side will tilt the platform pricing structure in favor of the other side of the platform. Complementor multi-homing reduces the net benefits complementors can obtain from platform network externality (Armstrong, 2006; Armstrong & Wright, 2007; Hagiu, 2009; Rochet & Tirole, 2003), because platforms would have monopolistic power only to multihoming complementors who seek to gain access to users single-homing on the platforms. Conversely, multihoming users will be beneficial to complementors and induce complementor exclusivity and increased investment to the focal platform (Athey, Calvano, & Gans, 2011; Choi, 2010).

Critique

In this section, I reviewed the fundamental concepts and predictions of platforms. In sum, technology platforms are conceptualized and modeled as a distinctive governance mode, which differs from traditional markets for innovation and technologies in its significant direct and indirect network externalities. The monopolistic market structure of winner-take-all is the most fundamental prediction as a result of such network externalities on platform-based innovation (Caillaud & Jullien, 2003; Katz & Shapiro, 1986; Lee et al., 2006). As complementors benefit from the increase of both users and complementors and vice versa, and such positive feedback loop also influences expectation of potential complementors and user expectations. In an equilibrium state, a dominant technology platform should retain all complementors and attract all user adoptions (McIntyre & Srinivasan, 2017; Zhu & Iansiti, 2012). While the literature on platforms is extensive, I only reviewed several branches that are most relevant to open source innovation, including the performance implication of platforms (WTA), platform entry and governance. From this most relevant literature, several research opportunities can be identified.

Research opportunity 1: platform-based technology competition is not fully addressed. One limitation of current research is loss of technological focus in most of the studies. While the concept of platforms was initially brought up by Katz and Shapiro (1986) to study technology adoption and diffusion, the majority of the studies on platforms focuses on prices and product competition, with the exception of a few studies on technology platform governance. In addition, in terms of methodology, most of the works are formal modeling. and the limited number of empirical papers focus almost exclusively on the video game industry. Therefore, we know little about the how the findings can be generalized into other settings, where unique boundary conditions may exist.

Research opportunity 2: platform-based nature of open innovation is not explicitly discussed in the current literature. Related to open source and open innovation, some foci of the platform literature display considerable similarity and overlap with the open source literature. For example, the governance and control is also a central topic in open source. However, the platform literature is distinct in that it focuses more on pricing and market competition, in addition to the innovation activities underlying the product on platforms. Also, in open source literature, the concept of platform is weak, without clear distinction of platform owners and core technologies from sellers and complements. In fact, in open source literature, the discussion of price is completely absent by nature.

Then, is open source platform-based innovation? The above reveal seems to give a definite answer. To a large extent, open source constitutes a two-sided market for technologies, which contributors provide technological input on the one side and users adopt the technologies to meet their own demand. Moreover, the network externality is distinctive in open source. Accordingly, open source technologies should display a tendency of WTA, such as Linux, Mozilla, as well as more recent Tensorflow. Conceptualizing open source as innovation platforms, however, also brings challenges unresolved by existing literature. First, open source lacks the price mechanism, which renders most of the discussion in the platform literature regarding entrance and price structure irrelevant. Yet, interestingly, we still see the emergence of WTA in some open source domains. Then, how does WTA emerge when there is not economic cost of switching and adaptation for users, when information is fully accessible to the public? In particular, is it possible that knowledge structure of the platform, jointly impacted by platform owners and complementors, can replace price structure in regulating open source platforms and contribute to the WTA situation?

The second related challenge is to define who is the complementor and platform owner in the case of open source. In appearance, defining the open source as platform-based innovation assumes contributors are the complementors/suppliers on open source as platforms. However, contributors, although providing knowledge supply, does not provide independent products in addition to the core technologies of an open source project. Rather, open source as platform-based innovation could be at a higher level, in the sense that each platform is composed of multiple open source projects that share the similar knowledge base or development framework. In the context of software development, such as shared knowledge base, constituting to the platform core technology, could be open standards such as programming languages, or developing environment or operating systems such as Linux. In this case, each open source project that is embedded in the developing environment becomes complementor to a platform.

To some extent, such definition of the boundary between platforms and complementors makes complementor particularly important. As the developing environment is loosely defined, and most often light weighted on purpose to allow higher flexibility, the extent to which the additional functionality provided by complementors becomes more critical in the adoption of platforms and diffusion of technologies. In the following section, I review the strategies that complementor can leverage to complete in platform-based innovation, with a particular focus on their expansion strategy on multiple platforms.

Research opportunity 3: few studies have investigated the impact of multihoming on individual complementor themselves. Despite the general conclusion that the overall complementor surplus reduces if all complementors participate in multiple platforms, we know little about the consequence of multihoming to specific complementors when the multihoming decision can vary across complementors. Most of the current analysis on multihoming assumed away the heterogeneity of complementors, regarding that supply by complementors as the qualitatively the same and accordingly user demand is the same for all complements. Such assumption is in contrast with the notion that network externality could be localized within platforms and the network structure, in addition to network size, could determine the extent to which a complementor in a given position can benefit from the network effect of platforms (Afuah, 2013; Lee et al., 2006; Suarez, 2005). One exception is Rasch (2007), in which the author noticed such negative effect of multihoming on complementors is weaker if there exist differentiation among complements, leading to partial multihoming of one side even at the equilibrium state. While such possibility is backed by the emerging empirical evidence that multihoming is related to the heterogenous perceptions and current performance of complementors (Bresnahan, Orsini, & Yin, 2015; Gu, Oh, & Wang, 2016; Hyrynsalmi, Suominen, & Mäntymäki, 2015), current research has not discussed how partial multihoming impacts may vary for those who initiated such strategy, and for their single-homing counterparts.

Relatedly, existing research on multihoming has yet considered is the interdependence among complementors. The current models of multihoming have not discussed such situation, and most of the empirical analysis focuses on platforms where interdependence is relatively moderate (such as the video game consoles). However, in complex platform-based technology system, such as telecommunication (e.g., Ranganathan & Rosenkopf, 2014; Toh & Miller, 2017), energy storage (Adner & Kapoor, 2010) and software applications (Kapoor & Agarwal, 2017), knowledge and technology interdependence of complementary innovation is essential, as complementors coordinate with and draws on knowledge from each other to satisfy user demands and deliver value (Kapoor & Lee, 2013). In the following section, we discuss the emerging growing on technology platforms as ecosystems that highlights the importance of technological interdependence. In our theory development, we combine both the research on multihoming on platforms as two-sided markets and innovation on platforms as ecosystems to study the impact of multihoming on complementors on its original platform.

Research opportunity 4: platform dynamics without price mechanism requires better understanding. Lastly, in studying both platform owners and complementors, the existing literature largely focuses on pricing strategies and its consequences. Although research on platforms as two-sided markets noticed the possibility of platforms with free complements (e.g., Boudreau & Jeppesen, 2015; Parker & Alstyne, 2005; Rochet & Tirole, 2003), unpaid complementors are mostly studied in the case of platform owner's subsidies (Parker & Alstyne, 2005) or competition with priced complementors (Boudreau & Jeppesen, 2015). The possibility that platforms can be organized without a price mechanism is essentially missing in the current conversation of multihoming. However, with the increasing popularity of open source and open technology standard, platforms that are essentially composed of free complementors who seek to profit not directly from their technologies and innovation, but from business model innovation (Teece, 2007), has become increasingly prevalent. As those platforms are organized without gatekeeping strategy and mandated exclusivity constraint (Boudreau, 2010; Caillaud & Jullien, 2003; Cennamo & Santalo, 2013; Parker & Alstyne, 2017), complementor multihoming constitutes a critical strategy that requires a deeper understanding in those open innovation platforms without a price mechanism.

INNOVATION ECOSYSTEM

The last literature relevant to the discussion on technology ecosystems. Compared with the economics roots in platform related research, technology ecosystem literature is largely derived from management and technology innovation that emphasizes on networks and interdependencies (McIntyre & Srinivasan, 2017). To some degree, research on technology ecosystems and two-sided markets overlaps, especially in the emphasis on the platform owner as a technological and market intermediate and on the critical role of complements and complementary technologies (Kapoor & Lee, 2013). However, the literature of ecosystems differs from that on two-sided markets in its emphasis on the supply side of technology platforms and the high level of coordination and the resulting technological inter-dependency of actors on platform ecosystems. Also, while platforms and two-sided markets emphasize the shared identities between complementors and owners, the ecosystems perspective does not assume a unified identity in coordinating innovation and transaction activities (Adner, 2017).

While the research on ecosystems in management research can be connected to an extensive literature studying technology evolution and inter-firm relation (e.g., Ahuja, 2000; Anderson & Tushman, 1990; Dosi, 1982; Gulati, 1998; Nelson & Winter, 1982), in this proposal, I focus on two specific topics that are most relevant to understanding open source innovation. The first is the growing literature on technology ecosystems (e.g., Adner & Kapoor, 2016b; Kapoor & Agarwal, 2017; Kapoor & Lee, 2013). The second related literature is the studies on technology committees for industry standard setting (e.g., Ranganathan, Ghosh, & Rosenkopf, 2018; Ranganathan & Rosenkopf, 2014; Rosenkopf, Metiu, & George, 2001; Toh & Miller, 2017).

Technology interdependencies in ecosystems

Compared with the literature on multihoming on platforms as two-sided markets that largely focuses on the market and price dynamics created by technology platforms, the ecosystem perspective emphasizes more on the technological supply on platforms and highlights the necessity of coordination and collaboration among suppliers within platforms. Accordingly, a variety of definitions of technology ecosystems emphasize the distinct positioning of complementors within the structure of a platform and how they interact with the central actor (platform owners) to jointly deliver the final set of technologies and value to meet the demand of users (Adner, 2017; Adner & Kapoor, 2010).

Under such emphasis on coordination, platform-based technology ecosystem is characterized by high interdependencies between platform owners and complementors (Adner and Kapoor, 2010, 2016; Gawer and Henderson, 2007). Complementary innovations are developed based on the technical features and standards of the core technology of the platform, as modularized extensions with additional functions that enhance the applicability and performance of the core technology (Toh & Miller, 2017). At the same time, the extent to which platforms can gain competitive advantage also depends on the capabilities and challenges of complementors (Adner & Kapoor, 2010; Afuah, 2001). Adner and Kapoor (2010), for example, found that the technological difficulties facing complements can reduce the competitive advantage of leaders of core technology who entered early global semiconductor lithography equipment industry. Relatedly, the availability of complementors can also affect the entry of potential competitors (Kapoor & Furr, 2015). The interdependence between core technologies and platform complementors is also reflected in the role of complementors in facilitating core technology innovation. Based on the data of medical device industry, Kapoor and Lee (2013) found that close collaboration with complementors through alliances increases a firm’s innovation investment to related technologies. In a more recent study, Toh and Miller (2017) found that the disclosure of core technologies can also be affected by the availability of complementors within an ecosystem in the telecommunication industry.

High interdependencies also exist among complementors. The use and changes of complementary technologies may require the presence and corresponding changes of others (Adner & Kapoor, 2010; Afuah, 2001), or they can be jointly used at the same time for better performance. Kapoor and Agarwal (2017), for example, elaborated such interdependencies of complementary technologies in the context of mobile application operating systems (iOS vs Android). In their qualitative interviews, they documented how software developers need to attend to the modification of the operating systems when developing their own applications so as to make sure the compatibility. Such interdependencies with other complementary technologies not only define the scope and performance of a complementor's innovation, but also influence the extent to which a complementor can benefit from the growth of the platform ecosystem and related network effects (Afuah, 2013).

The governance of ecosystems

Another important theme in this stream of literature is the governance of ecosystems. More specifically, research has been interested in the question, how firms can best navigate and manager such technology systems with high interdependencies. Existing literature has investigated two particularly critical governance mechanisms of technology ecosystems, as more specifically ways to manage those interdependencies, namely, technical committees and standard-setting organizations. In the following section, I will review and summarize the research related this theme.

Technology standards are prevalent in many high technologies settings (e.g., Ranganathan et al., 2018; Ranganathan & Rosenkopf, 2014; Rosenkopf et al., 2001). The standardization of products increases the compatibility. Hence, it allows both users and manufactures to enjoy a more extended network effect, knowledge spillover and economies of scope (Corts & Lederman, 2008; Farrell & Klemperer, 2007). Yet, the standardization beyond firm boundaries is difficult, because firms that can achieve standardization also compete with each other on the same market. The formation of technology standard committee provides an opportunity for firms to coordinate each other and to exert influence in the emergence of the dominant design (Ranganathan & Rosenkopf, 2014; Toh & Miller, 2017). Because of so, committees often play a critical role in shaping the development of technologies, in a way that is similar to the leader in the ecosystem. Meanwhile, the literature is also relatively independent from the ecosystem research as such concept is only discussed implicitly. Rather, to some degree, the research on technology committee is a diverse literature that incorporates literature on ecosystems, platforms and open source.

At the same time, early works on technology standards and committee are actually more closely connected with the platform literature. In Katz and Shapiro (1986), they explicitly discussed how a technology or market display strong network externalities, having compatible products would bring higher benefits to all participants. They also point out that such compatibility and be achieved by having industrywide standards. Through modeling, Jullien (2001) argues that one of the motivations of industry standards is that the competition brought by new entrants in platform-based markets will lower the profits for incumbents if price discrimination is feasible, forcing incumbent platforms to favor cross-platform compatibility and standardization. In those discussions, the concept of compatibility is often discussed as a part of platform strategies that allows multihoming (Rochet & Tirole, 2003). However, those research rarely addressed how the compatibility cross products and platforms through technology standard setting can be achieved in the first place.

Literature has also discussed how standard setting is achieved through coordination and communication among members. Studies taking this view argue that standard committee allows firms to "divergent views and interests, serving as loci for consensus-building and adjudication by bringing representatives from various organizations and coalitions together to define technological outcomes" (Rosenkopf & Tushman, 1998: 8). Hence, the formation of such committee to coordinate with players within an ecosystem is more likely to emerge after technological changes and before the emergence of dominant design (Rosenkopf & Tushman, 1998). The coordination process is in essence related to the network position of each actor. For example, based on a natural experiment of the 3G standard setting committee, Leiponen (2008) found that members' ties with other members within the information and formal standard setting give them more influence in the process of negotiating standards. Dokko and Rosenkopf (2010) also propose that another way for firms to gain influence is to hire individuals from other members, because such mobility can generate more social capital from the hiring personnel that allows firms to increase their influence in such settings. Studies have also examined the conflicts in standard setting. For example, Simcoe Rysman and Simcoe (2008) showed that distributional conflicts can cause coordination delays. Based on data from an important internet standard setting committee, the Internet Engineering Task Force, they measure distributional conflicts through the email content on committee's email lists. Also focusing on conflicts, Ranganathan and Rosenkopf (2014) found that central players in the technology less likely to oppose the current standard in the standard setting in the communication industry, as they possess more knowledge and capabilities in favor of existing structure and rules in line with the standard. However, if firms possess a central position in the commercialization network, they are more incentivized to promote change, as the existing widely adopted technologies can intensify competition in the downstream and hence the profit of those firms. Studying the same industry, Toh and Miller (2017) also found that the extent to which central player disclose information in standard setting is related to the extent to which they possess complementary technologies, because disclosure is less likely to induce competition when the leader also possesses unique complementary technologies and assets.

In terms of consequence, Rosenkopf et al. (2001) found that standard setting allows a higher rate of alliances formation, as the discussions and communications through such activities reduced perceived uncertainties, develop trust and embedded ties and shared knowledge of problem-solving. Waguespack and Fleming (2009) found that participating in open standard setting increases new ventures' likelihood of liquidation event in the context of computer software industry, because such participation not only allows new ventures to learn and gain more technological knowledge for better innovation, but also send out signals about their reputation and capabilities.

Critique

In this section, I reviewed the literature on technology ecosystems, which can also be applied to understanding open source platforms. More specifically, the literature can be divided into two substreams, with one explicitly focus on ecosystem interdependencies, and the other focuses on the ecosystem governance dynamics. Both pockets of research highlight the interdependency and coordination beyond firm boundaries in innovation and technology competition. The literature on the governance of ecosystems, such as standard setting organizations, provides rich details on the dynamics within the standard setting committee and outcome of such industry-wide standardization.

Research opportunity 1: performance of complementors in the competition of multiple ecosystems. Similar to the research on platforms as two-sided markets, most of the current studies on platform ecosystem still focus on the implications for owners of core platform technologies. With the exception of Kapoor and Agarwal (2017), little is known regarding how interdependencies and ecosystem structure may affect complementors, despite their critical role in platform ecosystems. Moreover, the current literature on platforms as technology ecosystems has largely focused on the within platform dynamics. The possibility of ecosystem strategies, in terms of both the leader and the complementor, can be shaped in the presence of multiple and potentially competing platforms, is yet considered in the current conceptualization of technology ecosystems. Yet, as competing ecosystems emerge, such consideration becomes increasingly important different players may face diverse incentives, either to join multiple ecosystems and gain additional value, or to reduce the negative impact of rivalry and competition (for platform owners).

At the same time, this literature focuses on the development of knowledge-based and coordination responsibility of a core technology is defined within a technology ecosystem by the power structure and negotiation processes. Few studies have investigated how participants adjust and adapt their technologies and capability development during competition once a consensus industry standard emerges. Another direction that may require further investigation is the competition among standards. Despite the frequently observed rivalry (e.g., HDDVD/ Blueray, CDMA/GSM), surprisingly few studies explored research question related to such phenomenon, especially with regard to how members, especially complementors, choose among standards and the factors that contribute to the competition outcomes. Even less attention is paid to how complementors' choice of the standard can shape their own innovation processes and outcomes. Lastly, the literature on ecosystems assumes proprietary rights of the technologies developed based on the standards, although the standard itself can be open (Waguespack & Fleming, 2009). Research has yet investigated how the open source nature of subsequent innovation can shape the standard setting choice and processes.

Research opportunity 2: demand-side dynamics in ecosystem competition. Unlike the platform literature, the demand/consumer side is rarely investigated in the current literature on ecosystems. Innovation is largely regarded as a supply-push process. The possibility that the formation and development of ecosystem and standard setting are influenced by firm's effort to curb the market demand is rarely investigated. Relatedly, few studies have considered the heterogeneity of consumers may play a role in determining how value is delivered through ecosystem coordination and standards setting for different participants and technology suppliers. One potential opportunity, as addressed in this dissertation, is to examine how consumers respond to the competitive dynamics of technologies ecosystems. Such investigation on the demand side would advance the understanding of venture performance within technology ecosystems.

SUMMARY

In the literature review section, I summarized the current state of development in three distinct literature related to open source innovation, (1) the studies on open source software (2) research on technology platforms (3) research on technology ecosystems and standard-setting committees. All three literatures are critically relevant to understanding the competition of technology as a system beyond firm boundaries. However, each of them is with a different focus. The studies on open source software largely regard such technologies as a phenomenon as a distinct innovation mode outside business world that does not concern value capture. The literature on technology platforms emphasizes the network externality and distinguish three types of players in the system, the platform, supplier (seller/complementor), and consumer (user). It also focuses more on the price and market dynamics, with the implicit assumption of independence among the suppliers/consumers. Technology ecosystem highlights the role of structure and coordination within an innovation system that shapes the value chain of an industry (e.g., Adner & Kapoor, 2010). At the same time, it also categorizes actors into ecosystem leaders and complementors. The demand side factor that contributes to such processes, however, is rarely addressed.

*** Insert Table 1.2 Here ***

Table 1.2 summarizes the critique of each literature reviewed in this section. Among the three streams of literature, only technology platforms touched among the inter-system competition. Apart from the literature on open source software, the technology platform and ecosystem literature both focus on proprietary technology systems with price signals. Those unaddressed possibilities highlight the potential contribution of the current study. Together, those unaddressed assumptions and questions shed lights on the need to systematically explore entrepreneurship in an innovation environment based on public knowledge with high interdependency in the absence of price signals and enforceable contract institutions. To understand an increasingly important phenomenon in contemporary innovation and entrepreneurship, this dissertation focuses on the impact of boundary decisions on venture performance in such environment. In the next two chapters, I first investigate the financial implications of boundary decisions for ventures' innovation activities, whether to internally develop technologies based on public knowledge or continuing seeking collaboration with external contributors or open source community. Then, I will explore how the choice of expanding to multiple open platforms, the boundary decision of a venture's product market, influences its existing customer base. By doing so, this dissertation seeks to advance the understanding of those important issues about innovation and entrepreneurship in open environment as summarized in this literature review.

Chapter II. Venture Growth and Expansion across Technology Platforms: Evidence from Open Source Platform Complementors

ABSTRACT

This study examines how a complementor's expansion to an alternative technology platform affects its user base in the original platform. Prior research has extensively examined the performance implications of the broadening of a firm's scope across industries. Yet, research is yet to examine whether existing insights apply to technology platforms, in which significant network effects exist and most providers of complementary products are entrepreneurs and small ventures that are resource-constrained. Strategy research on platforms, in turn, has highlighted the performance consequences of technological interdependencies between firms that create complementary products within a platform but has stopped short of investigating dynamics that unfold across platforms. We argue that a complementor's expansion to an alternative technology platform has a positive effect on its user base in the original platform as a result of inter-platform transfer of network externality. We empirically test our arguments using data on 2 million software technologies in 34 open source software development platforms. In support of our core proposition, our difference-in-differences analysis shows that a complementor that expands to an alternative open source software platform experiences a greater increase in user base than a matching counterfactual complementor. We discuss implications for research on firm scope, platform-based competition, and open innovation.

INTRODUCTION

Strategy research has highlighted the role of corporate strategies in driving firm growth. Dating back to the Penrosian insight that a firm's possession of slack resources creates the impetus for firm growth (Penrose, 1959), strategy scholars have extensively examined how a firm's decision to broaden its scope across industries and the performance consequences of such diversification decisions (e.g., Markides & Williamson, 1994; Montgomery & Wernerfelt, 1988; Rumelt, 1984). The broadening of a firm's scope across industries remains a vibrant domain in strategy research (e.g., Feldman, 2015; Sakhartov, 2018; Wu, 2013; Zhou, 2011). Yet, despite the emphasis that strategy research has placed on a firm's expansion across industries, strategy scholars have thus far under-examined a firm's expansion across open innovation platforms. Such paucity of research on this topic sits in sharp contrast with the role that open innovation platforms have on the innovation activities of technology-based firms that drive much of the growth in the contemporary economy (Chesbrough, 2006; Colombo et al., 2014).

The increasing prevalence of platforms in a variety of industries has spurred strategy research on platforms. Given the network externalities characteristic of platform-based competition (Katz & Shapiro, 1986), strategy research in this domain has highlighted complementarities between technologies in a given platform. Some studies have emphasized the performance consequences of technological interdependencies between firms that create complementary products (Adner & Kapoor, 2010; Kapoor & Agarwal, 2017). Greater awareness of these interdependencies has, in turn, spurred research on mechanisms that help firms manage them, such as standard setting organizations (Ranganathan et al., 2018; Ranganathan & Rosenkopf, 2014; Toh & Miller, 2017). However, this emerging strategy research on platforms has largely focused on dynamics occurring within a particular platform. Lacking in this literature is a focus on dynamics unfolding across platforms and, more specifically, on a firm's decision to broaden its scope and offer complements in different platforms.

Examining the implications of a complementor's expansion across open innovation platforms not only fills a void in strategy research by directing attention to an important yet under-examined contemporary phenomenon, but it also holds the promise of generating new conceptual insights. In contrast with the emphasis that existing literature places on slack resources as drivers of growth, the vast majority of complementors in open innovation platforms are entrepreneurs and small ventures (Gruber & Henkel, 2006; Waguespack & Fleming, 2009; Wen et al., 2015) and, as such, they are typically resource-constrained. Further, unlike the different industries to which a diversified firm expands, different open innovation platforms toward which a complementor may decide to expand function as partial substitutes to each other (Von Krogh & Von Hippel, 2003, 2006) and, accordingly, such expansion does not necessarily benefit from scope and scale economies that accrue to diversified firms. Hence, investigating the implications a complementor's expansion across platforms will likely help us shed light on other considerations underlying a firm's broadening of its boundaries. This study explores these opportunities by examining the following research question: *How does a complementor's expansion to multiple open innovation platforms affect its performance in the original platform?*

Because a key driver of a firm's performance in a context characterized by network externalities is the size of its user base (e.g.,Katz & Shapiro, 1986; Shankar & Bayus, 2003; Suarez & Utterback, 1995), in investigating this question we consider the effects that such an expansion has on a complementor's user base in the original platform. The prevalence of entrepreneurs and small ventures among complementors in open innovation platforms (Gruber & Henkel, 2006; Waguespack & Fleming, 2009; Wen et al., 2015) suggests that by diverting resources and expanding to an additional platform, a complementor may weaken its position in the original platform. Competition between platforms may also result in erosion of a complementor's original user base. However, contrary to those natural extensions of the logic underlying prior research, we argue that a complementor's expansion to an alternative open innovation platform has a positive effect on its user base in the original platform. As we elaborate in the theory section, two factors contribute to such main effect –expansion across platforms mitigates uncertainty about a complementor's prospects and facilitates that complementor's exploration of new opportunities to refine its complements. We probe the logic underlying our core proposition by examining contingencies that shape the effect stemming from these two factors. More specifically, we argue that the effect of a complementor's expansion to an alternative open innovation platform in mitigating uncertainty is stronger when a larger number of users are aware of that complementor and yet refrain from using its complements. We also argue a complementor's expansion to an alternative open innovation platform generates fewer opportunities for refinement when its complements have high levels of technological interdependence with the original platform.

To test these predictions, we focus empirically on the context of open source software platforms. This setting exhibits several features that are relevant to this study. First, open source software platforms have become an important context in which firms engage in open innovation and an increasing number of private firms also participate in those platforms in their innovation activities (Boudreau, 2010; Fosfuri et al., 2008; Shah, 2006). Second, in contrast with platforms wherein the price mechanism serves an important function in governing platform-based competition, such as those in the video game industry (Zhu & Iansiti, 2012; Cennamo & Santalo, 2013) and in the market for applications on mobile phones (Boudreau, 2012; Kapoor & Agarwal, 2017), open source software platforms, by their very design, do not rely on such mechanism. Third, although incumbent firms also participate in open source software platform, the vast majority of complementors in those platforms are entrepreneurs and small ventures (Gruber & Henkel, 2006; Waguespack & Fleming, 2009; Wen et al., 2015). Finally, this context enables us to observe millions of software program libraries in more than 30 open source software platforms, thus providing significant empirical traction for our empirical analysis.

In testing our predictions, an empirical challenge is to account for the possibility that observations might not be randomly assigned to treatment condition (*i.e.*, expansion to an alternative platform), which raises concerns with unobserved heterogeneity and reverse causality (Holland, 1986). For example, unobserved factors may exist that at the same time affect whether or not a complementor expands to multiple platforms and help explain the extent to which it attracts users in the original platform. In addition, the expectation to increase its user base can induce a complementor to expand to an alternative platform. In other words, expansion to an alternative open innovation platform might relate positively to an increase in a complementor's user base in the original platform, as we predict, but without being a causal precursor to that outcome. Thus, it is important to distinguish between the component of the increase of the user base that is indeed attributable to a "treatment" effect (*i.e.*, expansion to an alternative platform) and the component that results from a "selection" effect (*i.e.*, complementor's decision to engage in such expansion). To do so, we adopt a difference-in-differences approach that compares the change in a complementor's user base in the original platform before and after its expansion to an alternative platform relative to analogous change observed in a matching counterfactual observation. As we detail in the methods section, we match each platform that expanded to an alternative open innovation platform to another complementor that is otherwise similar but that remained only in the original platform. Our difference-in-differences analysis reveals that a complementor's expansion to an alternative platform results in an increase in that complementor's user base in the original platform. In the discussion section, we elaborate on the implications of our findings for research on firm scope, platform-based competition, and open innovation.

THEORY DEVELOPMENT

Background literature on platforms

Literature in economics has approached technology platforms as a distinctive governance mode of innovation Katz and Shapiro (e.g., 1986, 1994). Competition based on technology platforms differs from traditional markets for innovation and technologies in its significant direct and indirect network effects (Katz & Shapiro, 1986; McIntyre & Srinivasan, 2017; Schilling, 2002a; Suarez, 2005; Zhu & Iansiti, 2012). Not only does the increase in users generate higher utility and value of the technology for other users, but it also benefits the complementors providing complementary technologies (Armstrong & Wright, 2007; Parker & Alstyne, 2005; Zhu & Iansiti, 2012).

This literature views complementors as independent suppliers of technologies that build on the core technologies of the platforms. A few studies have explored the possible implications of complementor's expansion across competing platforms (Armstrong & Wright, 2007; Carrillo & Tan, 2006; Landsman & Stremersch, 2011; Rasch, 2007), with a focus on the consequences of such strategies to platform owners. For example, using data from the video game industry, Cennamo and Santalo (2013) showed that complementors expansion has a positive impact on a focal platform's market share, because complementor's inter-platform expansion, as a result of non-exclusivity, attenuates adverse selection of complements in the original platform.

Although those studies mostly focus on proprietary platforms with price signals, this literature has long noticed the existence of platforms with free complements (e.g., Boudreau & Jeppesen, 2015; Parker & Alstyne, 2005; Rochet & Tirole, 2003). Such unpaid complementors are mostly studied in the case of platform owner's subsidies (Parker & Alstyne, 2005) or competition with priced complementors (Boudreau & Jeppesen, 2015). For example, in a recent study about complementors on video game platforms, Boudreau and Jeppesen (2015) discussed the impact of free complements on the overall growth of the platform. They found that the increase of free complements lowers the overall rates of innovation for platforms, because free complements can diminish their priced conterparts' incentives to further invest in innovation.

Despite the notion of free complements, the possibility that platforms can be organized without a price mechanism is rarely addressed. However, with the increasing popularity of open source and open technology standard, open innovation platforms with unpaid complementors have become increasingly prevalent. As those platforms are organized without gatekeeping strategy and exclusivity constraint (Boudreau, 2010; Caillaud & Jullien, 2003; Cennamo & Santalo, 2013; Parker & Alstyne, 2017), expansion across multiple platforms constitutes an important strategy that complementors can leverage during competition.

More recently, the strategy literature has also paid increasing attention to technology-based platforms. In contrast with the literature in economics that approaches complementors as independent player in a platform, strategy scholars have adopted an innovation ecosystem to examine platforms and emphasized the high level of coordination among players to manger the interdependencies that exist in a platform context (Adner, 2017; Adner & Kapoor, 2010). Some studies have underlined the interdependencies the arise between platform owners and complementors (Adner and Kapoor, 2010, 2016; Gawer and Henderson, 2007). The extent to which platforms can gain competitive advantage hence depends on their abilities to manage those interdependencies (Adner & Kapoor, 2010; Afuah, 2001). Adner and Kapoor (2010), for example, found that the technological difficulties facing complements can reduce the competitive advantage of leaders of core technology who entered early global semiconductor lithography equipment industry, because the delays in complements innovation allow rivals to have more time to catch up. The interdependence between core technologies and platform complementors is also reflected in the role of complementors in facilitating core technology innovation. Based on the data of medical device industry, Kapoor and Lee (2013) found that close collaboration with complementors through alliances increases a firm's innovation investment to related technologies because collaboration lowers the cost of utilizing new technologies. In a more recent study, Toh and Miller (2017) found that the increasing availability of complementors within platforms in the telecommunication industry negatively affects the disclosure of core technologies, due to the resulting expropriation concerns.

High interdependencies can also exist among complementors. The use and changes of complementary technologies may require the presence and corresponding changes of other complements (Adner & Kapoor, 2010; Afuah, 2001), or they can be jointly used at the same time for better performance. Kapoor and Agarwal (2017), for example, elaborated such interdependencies of complementary technologies in the context of mobile application operating systems (iOS vs Android). In their qualitative interviews, they documented how software developers need to attend to the modification of the operating systems when developing their own applications in order to ensure compatibility. Such interdependencies with other complementary technologies not only define the scope and performance of a complementor's innovation, but they also influence the extent to which a complementor can benefit from the growth of the platform ecosystem and related network effects (Afuah, 2013).

Although strategy research has made strides in elucidating how independencies shape dynamics that unfold within the context of a platform, it has stopped short of examining dynamics that occur across platforms. Examining issues that arise across platforms is important, given that multiple platforms can co-exist at the same time (Adner & Kapoor, 2010), creating incentives for complementors to capture more opportunities by entering additional platforms. In the following section, we examine the implications of complementor's decision to expand its boundaries across open source platforms, by having complements in multiple platforms.

Hypotheses

When examining implications that a complementor's expansion across multiple open innovation platforms carry for its user base, we consider three distinct types of benefits that can accrue to the users on the original platform. The first benefit is related to the enlarged network effects for users. Positive network effects have been long regarded as the fundamental premise underlying platform-based technologies (e.g., Farrell & Klemperer, 2007; Katz & Shapiro, 1986; Parker & Alstyne, 2005). In the context of technology platforms, the utility function of platforms and complements for users increases with the growth of both users (direct network effects) and complementors (indirect network effects) within a platform, and vice versa. The presence of network effects has been widely studied as a most critical drivers of platform competition (e.g., Armstrong & Wright, 2007; Cennamo & Santalo, 2013; Corts & Lederman, 2008; Eisenmann et al., 2011). Users prefer platforms with larger user base and more complements, as they enable more convenient interactions with other users and provide a larger choice set of complements for wider range of functionally. Similarly, complementors are also attracted to such platforms because of the higher potentials to gain more users. Such positive feedback loop often results in the winner-taker-all situation, in favor of the platform who first tips the critical mass of user base (Katz & Shapiro, 1986; Lee et al., 2006; Schilling, 2002b).

While most current discussion on network effects is at platform level, we posit that network effects also affect user adoption of complements. Especially when complementors provide compatible complements to alternative platforms, such strategy could alter network effects users on the original platforms expect to accrue from adopting the complement. Multi-homing first increases direct network effects for users, because compatible complements by the same complementor allows users to coordinate with a larger scope of other users on other platforms.

Multihoming expansion of complementors also attract more users through the spill over of indirect network effects from the entering platform to the original platform. Although network effects are mostly theorized as dynamics within a single platform, a few studies have found that network effects may transfer across competing platforms under the presence of multihoming. In the context of video game industry, Corts and Lederman (2008) found that the positive indirect network effects for user can migrate to competing video game hardware platforms, as complementors seek multi-homing expansion to reduce the fixed cost of developing video games. In turn, users benefit from the growth of complements in competing platforms. Vice versa, it is also possible for complementors to benefit from the user growth on alternative platforms under the presence of multihoming complements. In the context of open source platforms, where the cost incur to users is mostly related to learning, similar functionality and underlying function-specific knowledge provided by technologies from multihoming complementors allow users from other platforms to access the original platforms with relative ease. Users on those alternative platforms may seek to reduce the fixed cost of learning of complements and access to a broader range of complements by adopting the technologies provided by the complementors on the original platform. Hence, multihoming also allows complementors to attract users from other platforms as a result of the inter-platform spill-over of network effects.

The second mechanism we considering is the potential signaling effect of multihoming expansion. Multihoming expansion increases the perceived utility of complements for users on the original platform by reducing the perceived uncertainty of the complements. Despite the openness of the platforms and transparent knowledge structure of complements, users face considerable uncertainty about their quality and the likelihood that they will continue to support and refine the technologies that they make available in an open platform, and such uncertainty can result in market failure (Akerlof, 1970). While the full disclosure of technologies through open source can partly assuage uncertainty about the quality of a complementor's technologies, such disclosure does not necessarily inform users whether complementors are willing to maintain and improve those technologies in the future. Such uncertainty of whether the complementor is incentivized and able to maintain the development of the complements is particularly distinctive in the open source platforms. On open source platforms, complements are provided for free and the technological capability barrier to entry is low. Those attributes make it difficult for users to evaluate whether the complements can be properly maintained and improved in the future when deciding on which complements to adopt. Such information asymmetry is even further aggravated by the absence of price signals, which is usually the most credible signal to transfer information and resolve such asymmetry (Akerlof, 1970; Spence, 1974). Hence, in open source technology platforms, the information asymmetry, specifically lack of knowledge towards the future development and performance of complements, still constitutes a major obstacle that can suppress user adoption. In such situation, expansion to multiple platforms constitutes a signal that mitigates uncertainty about future refinements of a complementor's technologies that users face when deciding whether to use those technologies (Spence, 1974; Stiglitz, 1975). As users observe a complementor's expansion to multiple platforms, they can be more confident that the complementor possesses adequate technological knowledge to reinforce and expand the market of technology and is committed to doing so even beyond the focal platform. Second, complementor's expansion to multiple platforms sends out a credible signal of technological stability of the complement in the future. As the complementor expands to multiple platforms and exposed to more potential users, the users in the original platform expect the complementor to have stronger incentives to maintain and improve their technology because of the higher potential for a larger market share across several platforms. By doing so, the complementor can offset the information asymmetry concern of users due to the lack of price mechanisms on open platforms.

The third benefits that can increase user base for multihoming complementors is related to the potential increase of complements' quality as a result of multihoming. In addition to the signaling effect, expansion to multiple platforms facilitates that the complementor's exploration of new opportunities to refine its complements. One important purpose for open source for contemporary companies is to solicit collaboration with external contributors and access the distant knowledge from the crowd (Afuah & Tucci, 2012; Piezunka & Dahlander, 2015). Yet, existing literature also shows that oftentimes organizations may not be able to fully utilize the knowledge acquired from the crowd in open source as they tend to simplify and rationalize the filtering of unfamiliar knowledge in such distant search (e.g., Piezunka & Dahlander, 2015). Expansion to multiple platforms, in this case, can increase a complementor's ability to capture and absorb distant knowledge through open source, as it involves proactive learning and adaptation to different complementors and users in other platforms. As complementors learn, acquire experiences, and solicitate collaborations with external contributors in the new open source technology platforms, they assimilate and incorporate knowledge of new platforms into their own knowledge repositories. By doing so, it is more likely that they can transfer knowledge from other platforms into the original platform. Hence, expansion to multiple platforms allows complementors to break local search within the original platforms and to explore distant knowledge that they can recombine with their existing knowledge when creating subsequent innovation. Similarly, expansion to multiple platforms also allows complementors to gain experience with users with more heterogeneous demand, increasing their ability to recognize unfulfilled technological and market opportunities on the original platform.

The argument of increased knowledge through expansion to multiple platforms can be backed by the extended evidence on the impact of how exploration through alliance, acquisitions and diversification strategy can positively impact on firm innovation (e.g., Katila & Ahuja, 2002; Miller, 2006; Mowery, Oxley, & Silverman, 1996; Rosenkopf et al., 2001; Stuart, 2000). The expansion of knowledge base and recombination benefits of exploration across boundaries argued in the current literature also applies to complementor's expansion across multiple platforms. Moreover, in our context of open source platforms, there are reasons to suspect that the effect of boundary spanning on innovation performance is even more prominent, as it lacks the protection of intellectual property that may inhibit the learning and knowledge assimilation when crossing multiple platforms. In turn, the complement's user base on the original platform will increase as its quality improves. For those reasons, we hypothesize the following:

H1: A complementor's expansion to multiple open source platforms increases its user base in the original platform.

User awareness

Next, we explore the boundary conditions of the above predictions with respect to the characteristics of users and the complementor. The first factor we consider is the user awareness for complementors' technologies. We argue that the extent to which multihoming increases the user base of complementors on the original platform hinges on user awareness for complementors' technologies. The higher user awareness, the more a complementor will benefit from multihoming in growing the user base on the original platform.

First, user awareness of complementors' technology reflect higher potential of direct network effects for users. High user awareness indicates higher possibility that the complementor already attracts user interests beyond the focal platform. Once multihoming, such complementors with high user awareness are more likely to convert the interested users in the alternative platforms to actual users. In turn, the complements on the original platforms also become more attractive, as they allow users to interact with a wider range of audience who adopt complementors' in the alternative platforms. On the other hand, if user awareness is low, complementors' multihoming is less likely to attract users in the alternative platform, as they face more severe challenges of liability of newness. In such situation, users on the original platform are less sensitive to complementor multihoming, as such strategy without user awareness is less effective in extending the scope of interactions and coordination beyond the original platform.

Second, another major mechanism underlying the H1 is that complementor expansion to multiple platforms will lower the information asymmetry and hence perceived uncertainty of users on the original platform. If this mechanism is indeed driving the predicted user adoption after such strategy, we should see a stronger effect for complementary technologies where perceived technological uncertainty is a more prominent constraint of user adoption. If large number of users are aware of and show interest to technologies developed by the complementors it is more likely that adoption will increase once the perceived technological uncertainty is mitigated by expanding to multiple platforms. In open platforms, high user awareness suggests that complementors already attract considerable interested users with high potential of adoption, and technological uncertainty could be the last hurdle of actual usage. In such situation, the reduced technological uncertainty through the signaling effect of multihoming is more likely to be well received by this wider scope of interested users, catalyzing a higher number of adoptions. Conversely, when user awareness is high, complementors may face other challenges in attain user base. The signaling of stability is less important if there are fewer users that show interests in the technology in the first place.

Hence, the positive effect of complementor expansion to multiple platforms on user use should be stronger if there are high user interests in complementor technologies. These arguments lead to the following hypothesis:

H2: User awareness of a complementor's technologies in the original platform strengthens the positive effect of that complementor's expansion to multiple platforms on its user base in the original platform.

Technological interdependency

High technological interdependency has been regarded as the fundamental characteristics underlying platform based technological ecosystems (Adner & Kapoor, 2010; Clarysse, Wright, Bruneel, & Mahajan, 2014; Kapoor & Agarwal, 2017; Toh & Miller, 2017). Especially in the context of open source technology platforms, where knowledge creation is largely cumulative (Boudreau & Lakhani, 2015; Von Krogh et al., 2003; Von Krogh & Von Hippel, 2006), the role of technology intercedence could be more prominent in shaping the outcome of complementor and platform competitive strategies. We propose that the interdependency of complementors' technologies with the external knowledge of the original platform also affect the extent to which multihoming increases the user base for complementors.

An important argument underlying H1 is the ability of complementor expansion to multiple platforms to attract new users from other platforms through the indirect network benefits. More specifically, new users are motivated to initiate the adoption of the complements on the original platform to lower their fixed cost of learning. One boundary condition of this argument, then, is the extent to which cost of learning is reduced by the shared knowledge underlying complementors' technologies on different platforms (Farrell & Klemperer, 2007; McIntyre & Srinivasan, 2017; Zhu & Iansiti, 2012). If complementor technologies on the original platform are highly interdependent with other knowledge, users outside the original platform would still incur higher cost of learning, as the focal complementors' technology would be a smaller portion of knowledge that users need to learn to explore the choices on the original platform. In such situation, multihoming has limited effect in attracting new users, because users outside the original platform still have to overcome high hurdles of learning and adaptation to implement complementors' technologies in its original environment.

Moreover, complementors who heavily rely on platform-specific knowledge may experience lower innovation quality in the new platform, reducing the likelihood of additional adoption by users from outside the platform. High levels of technological interdependence with platform-specific knowledge induces compatibility issues that lower complementor's innovation quality in the new platform, as the performance and functionality of complementor technologies may not fit the new platform environment. Indeed, studying the video game industry, Cennamo et al. (2018) recently found that expansion to multiple platforms can lower the innovation quality for complements developed for those newly entered platform. Hence, if the complementor has high levels of technological interdependence in the original platform, the knowledge about other platforms gained through expansion may not be applicable to improving the quality of their complements on the original platform. In addition, knowledge interdependency may weaken the positive signal of expansion, as the potential incompatibility issues on the new platform may cause users to doubt the overall technological capability of the complementor (Cennamo et al., 2018). With that, we arrive at the following hypothesis:

H3: *Technological interdependence with the original platform weakens the positive effect of a complementor's expansion to multiple platforms on its user base in the original platform.*

Platform competition

In conceptualizing the impact of multihoming on the adoption of complementary technologies, we highlight the possibility that network effects may transfer across competing platforms. Multihoming complements allow users on the original platform to interact with additional users from the alternative platform, while they also attract new users from the alternative environment as they lower the cost of learning to adopt the original platform. In both cases, multihoming expands the network effects of complements beyond a single platform.

Following this mechanism, competitive dynamics between the original platform and the new platform the complementor seeks to enter should influence the extent to which multihoming can benefit the user base of the complementor on the original platform. We argue that the benefits of multihoming is weakened by the competitive advantage of the original platform for two considerations. When the original platform possesses relative advantage over the alternative platform, the additional users that extend the existing users' expected network effects become less important. Users on the original platform are more likely to value interactions with users within the platform as the platform is winning and more users will join the original platform (McIntyre & Srinivasan, 2017; Zhu & Iansiti, 2012). Such expectation renders interaction and coordination with user on alternative platform less valuable and necessary. Similarly, users from the alternative platforms are more likely to adopt the original platform given the relative competitive advantage, regardless of the multihoming behavior. While they are still more likely to adopt the multihoming complements because of lower cost of learning, the increase is new users is more severely capped by the smaller size of the platform.

Conversely, when complementors from a disadvantageous platform expands to a more dominant platform, we should expect the greater impact of multihoming on user base for the complementor on the original platform. In such situation, the existing users place higher value on the interaction and coordination with other platforms, which have more extended users and can generate higher network effects. Moreover, as the alternative platform possesses a higher number of users, the increase of new users can be more pronounced once the complementor expands to the alternative platform. The possibility that multihoming is more beneficial to complementors started from disadvantage platform is in line with the current conclusion on how multihoming shapes platform competition (e.g., Caillaud & Jullien, Armstrong 2006; Armstrong & Wright 2007). The conceptual modelling by Caillaud and Jullien (2003), for example, shows that exclusivity increases the likelihood of winner take all, suggesting that multihoming would intensified platform competition by allowing the weaker platform to bridge the cap of user base. Figure 1 provides a summary of all hypotheses.

H4: *The competitive advantage of the original platform relative to the alternative platform weakens the positive effect of a complementor's expansion to multiple platforms on its user base in the original platform.*

Figure 2.1 provides a summary of the hypotheses.

Insert Figure 2.1 Here

DATA AND METHODS

Empirical setting: open source platforms

Open source is formally defined as "a decentralized…development model that encourages open collaboration…with products such as source code, blueprints, and documentation freely available to the public" (Levine & Prietula, 2013:1415; Wikipedia, 2018b). Open source platforms differ fundamentally from other platforms in which they house technologies that are non-proprietary, developed and distributed for free in absence of market price, and with full disclosure of knowledge and innovation process (Bagozzi & Dholakia, 2006; Lerner & Tirole; Von Krogh & Von Hippel, 2003, 2006).

To date, open source technologies have profoundly shaped contemporary innovation and technological change in many ways. In several sectors fundamental to computer science technologies, such as security encryption, server, data analytics infrastructure and mobile operating system, innovations such as OpenSSL, R, and Linux-based technologies have taken expressive market shares, ranging from 30% to 90%, in 2017 (BlackDuck, 2015; Techfae, 2016). In terms of value creation, it is estimated that the use of open source technologies has created $3 to $5 trillion dollars of economic value since 2005 worldwide, which equals to 20% of the U.S. GDP (Mckinsey, 2013).

Open source technology platforms have several appealing features that are relevant to this study. First, complementors play an extremely important role in open source platforms relative to core technologies, making the study of complementor strategies particularly relevant. Although open source has been traditionally regarded as independent (Bonaccorsi & Rossi, 2003; Comino et al., 2007; Raymond, 2001), in recent years it has been increasingly populated by profit-seeking organizations, particularly small ventures and entrepreneurs (Octoverse, 2018). At the same time, partly due to the open source nature, the core technologies underlying those platforms are usually lightweight, which provides abundant technological and market opportunities for complementors. The existence of those opportunities motivates considerable entrepreneurs and new ventures to participate in those platforms to appropriate value through business model innovation (Teece, 2007; Wen et al., 2015), making the investigation of complementor strategies particularly relevant.

Further, the non-proprietary nature of open source platforms makes expansion to multiple platforms an important and feasible strategy for complementors. Meanwhile, precisely due to the open nature of such platforms, research is yet to examine the performance implications of such expansion. In contrast with platforms wherein the price mechanism serves an important function in governing platform-based competition, open source software platforms, by their very design, do not rely on such mechanism. This is thus an appropriate setting to explore other performance implications facing complementors that expand across platforms, such as the consequences that such broadening of their scope carries in terms of those complementors' user bases.

Data sources

Our primary data source is GitHub.com, currently the world's largest host of computer source codes for open source software programs. GitHub started as a web-based cloud storage site for computer codes written a through distributed version control tool called Git, which functions as a "content tracker" for source code files in software development. On its website, GitHub as officially defined as "web-based hosting service for version control using git" (GitHub, 2018). The source codes of open source technology written through Git are stored, maintained and updated in the form of "public repository" on GitHub. Apart from storing source codes, GitHub also provides discussion boards for each repository, where users and developers of the open source technologies can post questions and suggestion for improvement (as "issues"). Another critical function GitHub provides for the open source technology repository is the "pull-request". Through pull requests, external contributors can make request to the owner of open source technology to incorporate contributor's changes to the source codes, and owners or administers can review the changes submitted through pull request before decision on whether to accept or reject changes. Figure 2A presents a visual example of such open source GitHub repository.

Due to the importance of version control and discussion channels in open source technology development (Frederiksen & Rullani, 2016; Von Krogh et al., 2003), GitHub soon become a natural host of the majority of open source communities and became the largest host of open source communities in the world in 2011 (Wired, 2016). Up to 2017, there are more than 5.8 million active users, 331k organizations that stored more than 19.4 million of active repositories (open source technologies) on GitHub (Octoverse, 2018). The owner of open source technologies is composed predominantly small ventures in the initial stage and individual entrepreneurs and hobbyist programmers. However, it should be noted that more than 50% of the fortune 500 and 600 public companies also have public repositories on GitHub (Octoverse, 2018).

GitHub documents the all the activities in Git-based development process and compiles into the meta-data of the technology that can be publicly retrieved by the Application programming interface (API) service (Dabbish et al., 2012). Currently, there multiple data sources where such information can be extracted. The original information of each GitHub activities, including updates and changes, discussions and pull requests are recorded by GitHub API in the form of JSON (JavaScript Object Notation) file, which can be downloaded direction from GitHub. This study primarily relied on GitHub Archive, a website that downloads of the JSON files on development activities from GitHub API and compiles into datadumps on an hourly basis since 2012 (Grigorik, 2012). We supplement this data with another similar website GH Torrent, which further complies the real-time activity level JSON data into relational database at activity, project, individual and organizational level data (Gousios, 2013). Lastly, our third major data source is libraries.io, which maps technology-level information for GitHub repositories, including the distribution channels and platforms, open source license and most importantly, the prerequisite technologies of implementing the focal innovation store in GitHub (Nesbitt & Pompilio, 2016).

Open source computer program libraries

In this study, we focus on a subset of upstream software technologies that constitute open source platform complements, namely program libraries or packages ("libraries" for short, hereafter). In Wikipedia, libraries are formally defined as "reusable codes and routines in computer programming"(Wikipedia, 2018a).

Libraries are typical complements to upstream programming languages and frameworks. They are developed to extend the functionality of certain platform-based programming technologies in the upstream of software development. The current open source libraries cover an extensive range within the software technologies, ranging from secondary programming languages, such as Typescript based on JavaScript, data analytics structure, such as Pandas and Numpy (data-frame structure) based on Python platform, to technologies related to machining and artificial intelligence such as Keras, Theano and Spark (optimizing compiler for distributed computation) based on Python platform.

While storing the source code files in repositories on GitHub, most of the libraries are distributed through platform-based channels. Those platforms are called as package managers and are specifically dedicated to one corresponding core technology. Package manager is formally defined as "a collection of software tools that automates the process of installing, upgrading, configuring, and removing computer programs for a computer's operating system in a consistent manner" (Wikipedia, 2018c). Through those platforms, users can download and use those libraries for developing downstream technologies or new libraries for further extension of the core technology's functionality. To some extent, libraries providers on publishing platforms are open source equivalent to third-party applications providers on App Store (platform) for IOS (core technology). However, it should be noted that compared with platforms like App Stores and video game consoles, where free and priced complementors co-exists (Boudreau & Jeppesen, 2015), program libraries and the corresponding platforms are open source by nature and can be accessed for free, as its core technologies are most often developed based on open technology standards.

Table 2.1 provides more examples of libraries as complements based on core technologies on a variety of package manager platforms, and Figure 2.2 (A&B) and Figure2.3-2.7 depict the innovation process in open source software ecosystems and the relationship between several key concepts and the corresponding terms specific to this technology setting. Together, the core technology of programming framework or languages, libraries as complements for extended functions, and the package managers for the distribution of libraries constitute to the platform ecosystems of each programming technologies.

Insert Table 2.1, Figure 2.2A, 2.2B & 2.3-2.7 here

Expect for only a few libraries (like the typescript case given above), our data show that the majority of libraries (more than 95% in the sample) are developed by small ventures and entrepreneurs, making the study of libraries as complements to technology platform ecosystem particularly relevant to technology entrepreneurship. Moreover, as libraries and the corresponding platforms are open source by nature and can be accessed for free, as its core technologies are most often developed based on open technology standards. Such features make package manager platforms for software development libraries programs an ideal setting for studying complementor expansion in open innovation platforms without price mechanism.

Lastly, our focus on libraries as platform complements is also motivated by the empirical feasibility for accurately measuring user base. As libraries.io compiled detailed data on the usage of each library that stores their source codes on GitHub through what is called "dependency file" in each repository, a list of pre-request libraries needed for implementing the source code from GitHub repositories. For example, if a repository is developing a downstream technology that involves data manipulation using the program language of python, it will list the library "numpy", which provides the data manipulation function for python, as a dependency in the "dependency file". We will discuss the measure of adoption in detail in the follow section. In summary, this list allows us to map the usage in each repository with high accuracy.

Data collection

As previously mentioned, the data on open source libraries stored in GitHub can be accessed through a variety of online data services. For both databases, we rely on Google Bigquery, on which both databases are available, for extracting and processing the data into the targeted level of analysis format (at complementor-month level). As the initial GitHub data from GitHub Archive is over 4 T and libraries.io data over 8 G, and the dyadic information on joint library usage exceeds 12 billion records, we take advantage of Google Bigquery's ability to process large scale data in relational database to identify the initial sample we later use for propensity score matching, resulting in 1.69 million libraries by 262,195 complementors on libraries.io mapped to the source code repositories hosted on GitHub[3]. As the level of analysis of the study as at complementor-

[3] We rely on the web address of library's source code to identify complementor and to map the libarires.io data with GitHub data. For each library that provide source code on GitHub, a unique and dedicated website will be generated for its source code repository as Http://github.com/complementor name/technology name, such address will not change once generated and is included in libraries.io database and in all activities records on GitHub related to the repository. For example, for the library of deep learning by Keras in figure 1, its address is https://github.com/keras-team/keras. Keras, in this case is the complementor of interest.

month level, our initial sample before matching contains 9,496,673 complementor-month observation on 251,221 complementors on 34 package manager platforms from since 2012, when GitHub became the absolute dominant open source hosting site for open source libraries (Octoverse, 2018).

Measures

Dependent variable

We measure *User base* using data about "dependency" in GitHub repositories. Dependency is a technical term in software engendering is used to "refer when a piece of software relies on another one" (StackExchange, 2015). In other words, dependency can be regarded as a prerequisite of downstream program or code. A common "dependency" one often encounters is the operation system requirement for installing software (i.e. the installation of a software may require Windows 10). Similarly, libraries often become the dependency of downstream application level software or of other libraries, as the performance of the latter requires the simultaneous presence and use of the former. From the perspective of libraries, which are specifically developed for code reuse, becoming the "dependency" of other computer programs is the only way such technology can be adopted by users (mostly programmers developing their own software).

We measure user base through the number of "dependent repositories". That is, the number of GitHub repositories that specifies a library from a focal complementor as 'dependency', the prerequisite of user's code. We counted the repository as a "user" of complementor's library if is counted as a "dependent repository". Hence, the user base we empirically measure is at the technology level, rather than the firm level (each firm or individual can become multiple users of the same focal library if they use the library as dependencies in several of their own codes). Such specification of dependency in GitHub repositories is summarized in Libraries.io, as illustrated in

Figure 2. In the figure, the highlighted "3.7K" repositories reflect the how many computer codes uses Keras (a library for deep learning in python) during their development. In our analysis, the variable is calculated as the logarithm of the cumulative number of adopters of complementor's library program technologies on its original platforms (repositories that specify a library provided complementors through a given package manager) up to the focal month.

Independent variables

Cross-platform complementor is a dummy variable set to one if a complementor published its libraries on two package manager platforms during the period of observations. Considering the matching design we will discuss later, the variable indicates whether a complementor belongs to the treated group (as 1) or the control group (as 0).

After is a dummy variable set to one if by the focal month, the complementor has published libraries on an additional platform other than its original platform. We determine the time of expansion through the create time of a library's corresponding source code repositories, data that is obtained through GitHub, and regarded the creation month of the library published in the additional platform (rather than the original platform) as the first month of treatment, where the variable switches from zero to one for the treated complementors.

Watchers. We use the number of *Watchers,* people who hope to keep updated with the latest development in the source code repositories of complementor's libraries to measure user interest. The variable is calculated as the logarithm of number watchers of complementor's library source code up to the focal month.

Own dependencies. We measure a complementor's reliance on platform knowledge through the number of dependencies specified by the libraries' own manifest file. That is how many other libraries are needed for the complementor's libraries to function. As each library contains several versions, the dependency of the library is a time-variant variable calculated as the logarithm of number of dependencies in the most updated version of the library up to the focal month based on the dependency information on libraries.io.

Relative platform size. We rely on the number of libraries of each platform to measure platform advantage between a complementor's original platform and the alternative entering platform at dyadic level. The number of libraries reflect the scope of complements, which is the most important factor attracting users. The measure is calculated as the ratio of libraries in the original platform in the month before the focal observation to that in the entering platform at the same time.

Control variables

Models control for the possible influence from other characteristics of a library's usage and complementor decision of cross-platform expansion. First, models considered the effects from external source of knowledge through collaboration, whether welcoming and incorporating contributions from external contributors, through the variable of *Pull requests*, which denotes the number of external contribution submissions to a libraries' source file up to the focal month. Because decision of expansion can be affected by the inherent quality of libraries, which also affects users' decision of adoption, we control for overall technical performance of the innovation through the number of *Issues* associated with the library, appear in the 'issue' section of the library's GitHub page, which is most commonly used for reporting bugs and errors. Secondly, the replication and morphing of complementor technologies on the original platform can also affect a library's adoption while influencing complementors' action of cross-platform expansion. Hence, models included the logarithm of the number of *Forks*, in which developers copied the source file of the library to their own repositories. In addition, the frequency of development activities of a library also contributes to the stability of its performance and hence perceived risks of adoption, affecting complementor expansion and user base at the same time. We account for this possibility by including the number of *Commits*, total number of times the complementor modifies the source files of the library up to the month.

Lastly, complementors' experience with the original platform can also affect expansion and user base on its original platform at the same time. To address this concern, models control for the *Number of libraries* published on the original platforms to the focal month, as well as *Tenure*, number of days since the complementor created the first library that was updated to the original package manager platform. As our contingencies considers platform advantage through relative platform size, we also controlled for the size of original platform through *Platform libraries.* Models also included complementor fixed effect and month fixed effects to account for the time-variant trend in the population that can affect complementor expansion beyond the focal platform and user base in the original platforms. Table 2.2 provides a summary of measures used in the analysis.

*** Insert Table 2.2 Here***

Estimations strategy: Difference-in-differences approach

To derive causal inferences for the proposed hypotheses, we face several challenges of endogeneity concerns. It is possible that observations might not be randomly assigned to treatment condition (*i.e.*, expansion to an alternative platform), which raises concerns with unobserved heterogeneity and reverse causality (Holland, 1986). The inherent complementor capability and knowledge structure could drive both expansion to multiple platforms and user base. For example, high quality and highly innovative complementor are more likely to have the knowledge and capacity to expand, while developing highly popular libraries on the original platform that is extensively used. Such possibility causes potential concerns of omitted variable bias. In addition, the current user adoption of complementors' technologies in the existing platforms may affect complementor decision making to expand beyond the focal platform, as they shape the perceived technological opportunities and the need to expand market share across platform boundaries, leading to the concerns of reverse causality. Thus, it is important to distinguish between the component of the increase of the user base that is indeed attributable to a "treatment" effect (*i.e.*, expansion to an alternative platform) and the component that results from a "selection" effect (*i.e.*, complementor's decision to engage in such expansion). To do so, we adopt a difference-in-differences approach that compares the change in a complementor's user base in the original platform before and after its expansion to an alternative platform relative to analogous change observed in a matching counterfactual observation.

We take advantage of the rich data in the package library platform context and construct a matched sample through one on one propensity score matching. For each complementor that enters another platform, we identify an otherwise similar 'twin' complementor who only focus on the original platform based on the probability scores of getting the treatment (cross-platform complementor) predicted by probit model. As shown in the probit regression of Table 2.3, the matching considers all the control variable and the platform information, as well as pre-treatment adoption and the time when the complementor enter the first platform (by the creation time of corresponding repositories). The propensity scores matching results in 18,841out of 37,600 treated complementor at the month of treatment (cross-platform complementors at the beginning of treatment) matched to the 18,841 out of 216, 834 control complementors with 6,114,124 complementors-month observations. As Figure 2.5 and Table 2.4 shows, the imbalances of the sample without matching is quite distinctive (although they display reasonable common support), in a way that treated complementors are significantly higher user adoption and high innovation quality (fewer errors), while they also differ from the control group in terms of collaboration and experience at the time of treatment. After the matching, we notice that the imbalances dropped to insignificant levels for most of the variables, particularly the number of *Issues* as a reverse coded proxy for innovation quality. The pre-treatment user adoption becomes significantly higher for control groups, which makes our estimation even more conservative, as additional increase of user adoption for treated group (if any) is unlikely to be a result of natural growth following the existing growth trajectories relative to its counterfactual.

*** Insert Figure 2.5, Table 2.2, Table 2.3 and Table 2.4 Here ***

Based on the matched sample, we compare the differences in user base before and after the entry under a difference-in-differences framework (DiD) (Donald & Lang, 2007), based on the following equation.

$$User\ base_{it} = \beta_c Cross\_platform\ complementor_i + \beta_m After_{it} + \beta_{cm} Cross\_platform\ compelmentor \times After_{it} + BX_{it} + \vartheta Y + \beta_i + \varepsilon_{it} \text{ (Equation 1)}$$

In equation 1 above, β_c captures the extent to which complementor receives user adoption on the original platform differ from their respective counterfactuals, while β_m refers to changes in user base in months after expansion. The coefficient β_{cm} captures the DiD effect that is, the difference in user adoption pre and post treatment observed between cross-platform complementors and its counterfactuals to treatment. B is a vector of coefficients on control variables $X_{i,t}$, ϑ represents the vector of coefficients on month dummies. After adding the complementor fixed effects (δ V), the time-invariant dummy of *Cross-platform complementor* was removed, leading to the following equation:

$$User\ base_{it} = \beta_m After_{it} + \beta_{cm} Cross\_platform\ compelmentor \times After_{it} + BX_{it} + \vartheta Y + \beta_i + \varepsilon_{it} \text{ (Equation 2)}$$

For testing the interactions as proposed in H2 and H3, we used the following model, in which the variable $Contigency_{it}$ refers to the contingent variables in H2 and H3 respectively *Watchers* and *Own dependencies.*

$$\begin{aligned} User\ base_{it} = {} & \beta_m After_{it} + \beta_{cmb} Cross\_platfrom\ compelmentor \times After_{it} \times Contigency_{it} \\ & + \beta_{cm} Cross\_platform\ compelmentor \times After_{it} \\ & + \beta_{mb} Multihoming_{it} \times Contigency_{it} \\ & + \beta_{cb} Cross\_platform\ compelmentor \times Contigency_{it} + BX_{it} + \vartheta Y + \delta V \\ & + \beta_i + \varepsilon_{it} \text{(Equation 3)} \end{aligned}$$

RESULTS

Before discussing the analysis, we check whether the propensity score matching resulted in control complementors that are indeed similar to the treated complementors. The first column of table 2 reports the mean of criteria for matching, including control variables in the overall sample before matching. Because all of the criteria for matching are stock measures of complementors on the original platform, we match the sample based on complementor performance and activities at the time of treated complementor's expansion. This causes the sample to include only the 1 observation of the treated complementor at the month of treatment and all month-complementor observation for the control group because they are all potential candidate as the counterfactual of the treated complement at treatment time. Hence, the presentation of treated complementor is drastically underestimated in column 1 of table 2. The actual percentage of treated complements take up to 25% of the entire population. After the matching, the sample reported in the last three columns of table 2includes 1 observation of each treated complementor at the month when it started treatment and 1 observation from the control complementor that resemble the treated complementor at a similar time.

As the last three columns of Table 2.4 shows, the *t*-statistics for control variables after matching was drastically reduced, and there are no significant differences between the treated and control variables in terms of several critical control variables, including number of *Watchers*, *Pull requests*, *Issues* and complementor *Tenure* on the platform. We further conducted graph analysis to examine whether the matching establishes a similar trend pre-treatment between treated and control observations as a baseline for DiD models. Figure 2.6 plots the user base of treated and control 6 months before and after treatment. It shows that the growth of users between the treated complementors and control complementors displayed very similar pattern before the treated complementors started treatment. However, after treatment, the treated complementors experience more drastic growth in user adoption and the gap of users between the treated and control complementor widen over time. The evidence suggests preliminary support for H1, which argues that expansion to multiple platforms increases user base of technologies developed by such complementors.

Insert Figure 2.6, Table 2.5 and Table 2.6 here

Table 2.5 summaries the statistics of mean and standard deviation and correlation matrix of the matched sample. Table 2.6 presents analysis based on OLS estimations for the expansion effects on complementor specific user base. Models 1-3 are models without month-fixed effects and complementor-fixed effects. In Model 1, only control variables are included. Model 2 adds the variables *Cross-platform complementor* and *After*, and model 3 includes the DiD coefficient of *Cross-platform complementor X After*. Model 4-6 presents equivalent models with month-fixed effects and complementor-fixed effects. Because those models contain complementor-fixed effects, the variable *Cross-platform complementor*, which is a time-invariant indicator of the treated are removed from the model. In all models, the variables used in interaction terms are mean-centered and the VIF is below the threshold of 10 and even the more stringent threshold of 5 (Kleinbaum, Kupper, Muller, & Nizam, 1988), indicating that the models do not pose significant concerns of multi-collinearity despite the high R-squared in Model 4-6 after adding complementor-fixed effects. We based our analysis on the final model with all fixed effects (Model 6 in Table 2.6).

H1 predicts that expansion to multiple platforms increase complementor specific user adoption. Consistent with this argument and with previous graph analysis, the coefficient on *Cross-platform complementor X After* is positively significant ($\beta = 0.203$, $p < 0.001$) in Model 6). The results indicate that holding other variables constant, expansion to multiple ecosystems increases user adoption for complementors by 20.3%, providing strong supports to H1.

*** Insert Table 2.7 here***

H2 argues that the positive effect of expansion to multiple platform ecosystems is stronger for complementors with extensive user awareness in their technologies on the original platform. To test the contingent effect, we used three-way interactions based on Equation 3 and the results are reported in Model 1 of Table 2.7. In line with argument of H2, Model 1 in Table 2.7 shows that coefficient on the interaction term of *Cross-platform complementor X After X Watchers* is significantly positive ($\beta = 0.046$, $p < 0.001$), revealing that the positive effect of expansion to other platforms is even stronger if complementor can attract extensive awareness in the original platform.

H3 predicts that the positive effect of complementor expansion to multiple platforms is attenuated by complementors' reliance on platform-specific knowledge reflected through the number of *Own dependencies* required for their own technologies complementors. Similar to the test of H2, the effect is examined through the three-way interactions based on Equation 3. The results in Model 2 of Table 2.7 are consistent with H3. More specifically, coefficient of *Cross-platform complementor X After X Own dependencies* is negative and significant (β = - 0.030, $p < 0.001$), showing that the positive effect of such expansion is weaker for complementors that have high interdependency with the focal platform. In sum, the analysis provides strong support to H3.

H4 argues that the positive effect of complementor expansion to multiple platforms is attenuated by the relative advantage of the original platform. The coefficient of *Cross-platform complementor X After X Own dependencies* is negative and significant (β = - 0.002, $p < 0.001$), which supports H4 and shows that the positive effect of such expansion is weaker for complementors that have high interdependency with the focal platform. Hence, H4 is support.

Robustness checks

We further conducted several sensitivity checks to test the robustness of the results. First, to rule out the possibility that the results are sensitive to the specification of time window (the 6-month window before and after treatment), we analyzed data of user adoptions based on the same model specifications in Table 5 and Table 6 using alternative time windows. In Table 2.8, Model 1 – model 3 are results using a 3-month time window, and model 4-6 are results based on a 12-month time window pre and post the treatment of treated complementors. Those results are highly consistent with the main analyses, providing additional support to the robustness of our conclusion.

*** Insert Table 2.8 and Table 2.9 here***

Another potential concern is related to the skewness of user adoptions. As revealed in the summary statistics of Table 2.5, the user adoption, as well as some of the control variables are with high standard deviations and relatively low means, indicating the data is possibly skewed. Similar to other technologies and business in general, the software libraries on platforms are characterized with a few highly successful and widely used libraries. At the same time, there is a considerable number of libraries are rarely used. While the skewness of the data may cause the results to be sensitive to outliers, our large same with over 10,000 libraries and over 300,000 observations should be able to mitigate this concern. To further alleviate such concern, we also tested models excluding outliers outside the 1% percent upper bound of user adoption, and the results remain fully robust.

The last threat related to the skewness of the data is that the inclusion of large number of low usage libraries undermines the validity of the results in applying to the relatively actively libraries that are truly important to platforms and to complementors. To address this concern, we constructed an alternative matched sample that only included treated complementors that had successfully attracted usage on the original platform before treatment for identifying matched pairs. This results in 9546 pairs of treated and control complementors after matching. As reported in Table 8, the analyses show similar effect of treatment for those more successful complementors. Holding other factors constant, expansion to multiple platforms causes nearly 20% increase in user base if the complementor has already obtained usage before treatment. The evidence provides further strengthens the robustness of the main analyses and findings.

DISCUSSION

This study examines implications of a complementor's decision to broaden its scope by expanding to multiple open innovation platforms. More specifically, this study proposes that a complementor's expansion to multiple open innovation platforms contributes to increasing its user base in the original platforms. Further, this study proposes that users' awareness about a complementor's technologies exacerbates this effect, whereas technological interdependencies with the original platform attenuate such effect. Our analysis accounts for the possibility that complementors are not randomly assigned to treatment condition (*i.e.*, expansion across open innovation platforms) which creates endogeneity concerns. To increase confidence that expansion across platforms is a causal precursor of the increase in a complementor's user base in the original platform, we adopt a difference-in-differences approach. In support of our main proposition, we found that a complementor's expansion to multiple platforms increases user adoption of that complementor's technologies on the original platform by over 16%. Findings also reveal that such effect is stronger for complementors that had attracted extensive user interest in their technologies but that had nevertheless refrained from using them. Finally, the analysis shows that such effect on a complementor's user base is weaker when the complementor maintains a high level of technological interdependencies in the original platform.

Limitations

One lingering question left is whether our findings about a complementor's expansion to multiple open source platforms is generalizable to other platform types. The lack of price mechanism in our setting contributes to the absence of platform control and entry barrier for complementors, which could affect the motivation of expansion for complementors in the first place. At the same time, free platforms and complements also affect the interdependencies of technologies and the cost of adoption facing users. On the one hand, it is possible that in priced platforms, such expansion effect on user base should still exist, if not stronger, as the signaling effect of cross-platform expansion in mitigating uncertainty is stronger when complementors face entry barrier. On the other hand, priced platform could also reduce the effect of other mechanisms driving the increased user adoption post-expansion, especially if price induces switching cost for users that inhibits user migration across competing platforms. One potential avenue for future research how free platforms and priced platforms may differ to gain a better understanding how the outcome of complementor strategies may differ under distinct forms of platform governance.

Theoretical contributions

This study advances the understanding of platform-based innovation first by considering the implications of inter-platform competition to complementors. In contrast with the emerging research that focuses on the within-platform dynamics, this study highlights how external competing platforms can also shape a complementor's competitive advantage within a platform ecosystem. More specifically, the hypotheses in this study reveal that the potential benefits of complementor expansion come not only from the potential gain of new user base in the entering platform, but also from a somewhat counter-intuitive increase of users in the original platform. This study also bears practical implications for the decision making of the complementor entrance. The boundary conditions discussed in this study suggest that entering multiple platforms is the most ideal for complementors that are central to the original platform while with a less embedded user base.

Secondly, this study adds to the research on the competitive strategies of platform owners. Consistent with prior research that suggests giving up exclusivity may benefit platform owners during competition (e.g., Boudreau, 2010; Cennamo & Santalo, 2013). While existing literature has discussed the potential concern of adverse selection in complementor exclusivity (Cennamo & Santalo, 2013), this provides another insight into the benefits of non-exclusivity. That is, the entrance to a competing platform may transfer network externality of the competing to the original platform through technologies provided by the entering complementor.

This study also contributes to the literature on open innovation. While existing research on open source focuses on the motivation of external contributors innovation (e.g., Belenzon &

Schankerman, 2015; Shah, 2006; Von Krogh et al., 2003) and comparison between open and closed innovation (e.g., Bonaccorsi & Rossi, 2003; Felin & Zenger, 2014), this study investigates competition strategies in open source in greater detail. Such shift of focus is particularly relevant given the increasing prevalence of open source among for-profit technology companies, who participate both as platform owners and complementors. As this study approach open source as technologies platforms without price mechanism, such investigation not only provides a new theoretical angle to understand competition in open source, but also provide managerial implications to the optimal strategy of managing and participating platform based open source innovation.

Chapter III. Can Free Resources Create Economic Value? Crowd Contributors and Venture Capital Investment to Open Source Technologies

ABSTRACT

Open source has become increasingly populated by entrepreneurial firms. Yet, it is unclear how open source-based high technology ventures can survive and sustain growth, when they completely disclose their knowledge and distribute innovation for free. In this study, we investigate how crowd collaborations with external contributors, an important and unique phenomenon in open source technologies, impact on the value creation of the inventing ventures as reflected in venture capital investment. In contrast with existing literature that regards the crowd as resources for innovation and inputs for knowledge creation through problem-solving, we highlight the value of crowd collaborators as rare and valuable market resources. For contributors in open source communities, the crowdsourcing collaboration is a sense-making process that allows them to develop a deeper understanding of firms' technologies. The familiarity with the underlying knowledge and shared identity developed through the collaboration with the crowd lock in the crowd as lead users who play a critical role in the diffusion of technologies in the product market. Hence, crowd collaboration reflects the inventing ventures' potential to profit from the user base, which will be reflected in venture capital investment as a major indicator of economic value for entrepreneurial ventures. We test our hypotheses using data on open source-based ventures' development activities from GitHub.com, under a matching design. This study contributes to the literature on collaborative innovation through crowdsourcing and open source by exploring its financial implication. Our focus on venture capital investment also deepens the understanding of how venture capitalists evaluate the economic value of open source innovation.

INTRODUCTION

Open source has become increasingly proliferated with entrepreneurial ventures (Alexy & Reitzig, 2013; Chesbrough, 2003; Von Krogh & Von Hippel, 2006). Although forfeiting the proprietary rights of technologies, open source can also bring unique benefits to new ventures' innovation and growth (Colombo et al., 2014). A distinctive feature of open source technologies is the extensive collaboration with external contributors in their corresponding online communities (Belenzon & Schankerman, 2015; Dahlander & Piezunka, 2014; Garriga, Von Krogh, & Spaeth, 2013; Lee & Cole, 2003). Existing literature have noted that, new ventures can access the human capital and knowledge inputs for free through collaborating with the crowd contributors in open source communities, as those external contributors volunteer to improve ventures' open source technologies without monetary compensation (Boudreau, 2012; Fleming & Waguespack, 2007; Lakhani & Von Hippel, 2003).

However, the economic implications of collaborating with such crowd contributors through the open source communities remain unclear for new ventures. As open source communities often lack the monetary reward (Nagaraj & Piezunka, 2017; Shah, 2006; West, 2003), knowledge from the crowd is free and publicly available (Lakhani & Von Hippel, 2003), which runs counter to a central tenet in strategy research that private knowledge is key to market competition and profiting from innovation (Barney, 1986; Dierickx & Cool, 1989; Eisenhardt & Martin, 2000; Peteraf, 1993). Even when collaborating with external contributors can bring knowledge benefits, the open source technologies resulting from such collaboration cannot directly capture economic value (Bloodgood, 2013; West & Gallagher, 2006). Ventures still need to search for appropriate business model or develop other proprietary technologies based on the capabilities gained in open source to materialize the value capture (Bonaccorsi et al., 2006; Casadesus -

Masanell & Zhu, 2013; West, 2003), making the economic promise of such free technological resource from contributors even more uncertain.

To date, the emerging literature on open source entrepreneurship has not addressed the puzzle, whether and how the collaboration to obtain free knowledge and technological resources can translate into financial benefits that ultimately sustain the growth of open source-based new ventures. In this study, we seek to explore this question by investigating the impact of collaboration with contributors on venture capital investment to open source-based ventures. Our focus on venture capital investment is rooted in its importance as an external financial resource for high technology new ventures (Barney, Busenitz, Fiet, & Moesel, 1996; Gompers & Lerner, 2001). At the same time, investment from venture capital also constitutes an important milestone that reflects a venture's economic value (Baum & Locke, 2004). Moreover, venture capital investment can be particularly crucial to open source-based ventures, as such ventures cannot directly generate profit by selling their innovation on the technology or product market, and hence are in more urgent need of external financial resources for subsequent development.

We propose that collaboration with the crowd increases the likelihood of venture capital investment to open source-based ventures, as the crowd functions as signals valuable market resources (Akerlof, 1970; Spence, 1974). In contrast with the existing literature that regards crowdsourcing collaboration as a process of knowledge creation in the upstream innovation process, our theory development the value of the crowds in the competition of the downstream product market. More specifically, crowdsourcing allows external contributors, who are oftentimes lead users that can profoundly shape the diffusion and adoption of new technologies (Rogers, 2010; Suarez, 2005), to develop a better understanding of the technologies. The familiarity with the functionalities, fundamental logic, and communication pattern, generated in the process of crowdsourcing, makes it easier for firms to lock-in critical users at the emergence of the downstream product market for the technologies. Hence, the crowdsourcing allows entrepreneurial firms to cultivate loyal users as rare and valuable market resources at an early stage. The resulting high potential of economic value capture will then be reflected through venture capital investment.

To test our hypotheses, we leverage a unique dataset documenting over 400,000 ventures centered on upstream software development technologies using a matching design, in which we use the time taken for crowd collaborators to respond to the venture as an instrument for collaboration completion. Preliminarily analyses show strong support for our arguments. We discuss the implications of our study at the end of this proposal.

THEORY DEVELOPMENT

Collaboration with the crowd in open source communities

Accompanied by its growing popularity in practice, collaboration with external contributors through crowdsourcing has gained increasing attention in the literature on innovation and strategic management. Crowdsourcing generally refers to the firms' behavior of soliciting suggestions and solutions to problems from the population of individual external contributors (Afuah & Tucci, 2012; Piezunka & Dahlander, 2015). In open source technologies, crowdsourcing is particularly relevant as it usually provides an online community for external developers and users to interact with each other and with the inventing venture. Such online communities become a platform for crowdsourcing collaborations, where actors jointly discover and solve technological problems while seeking to improve the innovation (Ebner, Leimeister, & Krcmar, 2009; Lee & Cole, 2003; O'Mahony, 2003). In those communities, the collaboration with external contributors in those communities oftentimes driven by the intrinsic interests of the crowd without monetary reward nor proprietary rights (Bagozzi & Dholakia, 2006; Waguespack & Fleming, 2009). Hence, those such external contributors in open source communities constitute free resources that firms can access through crowd collaboration.

Most of the current studies view such crowdsourcing collaboration with external contributors as a form of knowledge sourcing mode, which allows firms to search for distant knowledge that can facilitate knowledge creation and firms' adaption in the face of new technologies (Afuah & Tucci, 2012; Piezunka & Dahlander, 2015). While knowledge creation requires the recombination of novel and distant knowledge (Grant, 1996; Hargadon & Sutton, 1997; Katila & Ahuja, 2002; Nelson & Winter, 1982), firms tend to search locally for solutions (Levinthal, 1997; March, 1991), facing considerable hurdles to access and understand distant information and knowledge due to inertia , path dependencies and information filters, and limited absorptive capacity (Cohen & Levinthal, 1990; Nelson & Winter, 1982). As the crowd constitutes individuals with heterogeneous knowledge background, collaboration with the crowd allows firms to access a broader scope of knowledge to identify the "global optimal" solution. Indeed, empirical evidence has supported that collaborating with the crowd with more distant knowledge domains allows firms to identify high-quality solution during innovation (e.g., Jeppesen & Lakhani, 2010).

Existing literature also contends that crowdsourcing also aids the growth of the ecosystems backing a focal innovation developed by the firm. For example, Boudreau and Jeppesen (2015) noticed the existence of unpaid complementors in platform-based innovation, which is also in essence external crowd collaborators providing add-on solutions to a focal technology. Study 85 video game platforms, they found such collaboration with crowd increase the rate of innovation as platforms grow. The increasing availability of complementary technologies developed by the crowd collaborators then magnifies the network effects, expanding the focal innovation's advantages during technology competition (Adner & Kapoor, 2010; McIntyre & Srinivasan, 2017).

Recent literature, however, also uncovers the downsides of crowdsourcing in the process of knowledge creation. For example, Piezunka and Dahlander (2015) found that even though firms may access the distant knowledge from crowd collaboration, they may not be able to assimilate and utilize such knowledge as the filtering and learning process of such external knowledge is still subject to firms' own path dependencies and limited absorptive capacities. Moreover, collaborating with the crowd through open source communities also faces unique challenges. Knowledge from such external contributors in open source communities is often portrayed as atomistic, amorphous, with indistinguishable knowledge components (Bayus, 2013; Howe, 2008). For open source based entrepreneurial firms, it is also costly to develop routines to guide external collaborators and maintain communication (Dahlander & Frederiksen, 2012; Foss et al., 2016). The hazard of collaboration without well-defined contracts can be high in open source, as contributor participation in collaboration can be highly fluid. Those downsides bring doubts about how collaborating with the crowd can generate and capture value for entrepreneurial firms developing open source technologies.

Crowd collaboration and venture capital investment to open source technologies

In this study, we shift the focus of crowd collaboration to its impacts downstream on the product market to address the puzzle whether and how collaborating with the crowd creates value for open source technologies and their inventing firms. In investigating the mechanisms through which crowd collaborations generate economic value for ventures, we highlight the role of crowd collaboration in accessing gain users base on product markets, as critical prerequisites for ventures to subsequently profit from their open source technologies.

In the process of profiting from innovation, new ventures with open source technologies face a unique challenge. Unlike traditional proprietary innovation, for open source-based new ventures, innovation is separated from the actual value capture. The absence of prosperity rights makes it difficult for venture capitalists to gauge the potential economic value of the technology, even when the technical quality of an innovation can be verified. Open source-based ventures often need to resort to business model innovation or launching subsequent priced technologies, to capture the value of their open source technologies (Alexy & Reitzig, 2013; Fosfuri et al., 2008). However, despite the diversity of business models, such value capture process all requires an extended and loyal user base, with the potential to pay for service or priced alternatives. Hence, the fundamental role of the user base in such business model-based value capture process makes the competition of user and market share particularly pertinent for open source-based ventures.

We argue that collaborating with the crowd constitutes an efficient strategy to gain and reinforce the user base as a valuable resource for firms' technologies. The first type of market resources is the external contributors who directly participating in crowd collaboration. Existing literature on open source communities has noticed that most of the crowd contributors are themselves users of the open source technologies (Bagozzi & Dholakia, 2006; Baldwin & von Hippel, 2011; Chatterji & Fabrizio, 2014; Von Krogh et al., 2003). For those users, the collaboration constitutes a sense-making process that allows them to develop a better understanding of a focal technology. As they seek to contribute and improve the innovation, external contributors need to comprehend the principals, as well as the development logic and the underlying knowledge, above and beyond simply using the technology to participate in the development process. For one thing, collaborators are more likely lock-in to the technology, not only because the technology is modified by themselves to better suit their own needs, but also because they become increasingly familiarly with the technical details of the focal technology. At the same time, the crowd's familiarity with the focal technologies increases their switching cost to competing technologies, which also reinforce the unique user base of the technology. For the other thing, the collaboration also reinforces the social and emotional attachment of those users. As the community-based collaboration often requires frequent communication (Dahlander & Magnusson, 2005; Foss et al., 2016), during which the crowd develops shared routines and common experiences with the venture and other members within the community (O'Mahony & Ferraro, 2007; Shah, 2006). In turn, the shared identity and trust developed in such process also help ventures retain collaborator-user base.

Secondly, the crowd also facilitates the diffusion of technology to other users, further strengthening the firm's advantage in the market competition for users. As mentioned earlier, the crowd of users who seek collaboration are often lead users with considerable technological knowledge. Those lead users play a critical role in the diffusion of technologies, as other users tend to rely on their recommendation and education when deciding which technology to adopt (Rogers, 2010; Von Hippel, 1986). Hence, retaining the lead users through crowd collaboration allows the inventing firm to generate direct network effects among users to win over a larger proportion of market base(Cabral, 1990; Cennamo & Santalo, 2013). Such potential of capturing value from the user base, then, will be reflected in the valuation of the venture in the form of venture capital investment.

H1: Collaboration with crowd contributors increases the likelihood of receiving venture capital investment for an open source-based venture.

Although the crowd is often portraited as "free" and "unpaid" (Boudreau & Jeppesen, 2015; Lakhani & Von Hippel, 2003), attracting crowd collaboration is in fact not without cost.

The cost of crowd collaboration, then, may attenuate its value capture effect as reflected in venture capital investment. More specifically, we argue the major cost involved here is the opportunity cost of disclosing the technological components in open source communities – knowledge that ventures could have otherwise made proprietary. If crowd collaborations are based on extensive disclosure of ventures' own knowledge, the acquisition of user base through those external contributors can be more costly. Ventures may lose opportunities to directly capture value from such knowledge through intellectual property rights. In turn, the positive effect of collaborating with the crowd on increasing firm value would be attenuated during venture capital investment, due to such existence of opportunity cost.

H2: The venture's knowledge disclosure in the open source communities weakens the positive effect of collaborating with crowd contributors on the likelihood of receiving venture capital investment for an open source-based venture.

The second contingency we explore is related to the knowledge structure of the venture. The mechanism underlying the main hypothesis is that the collaborating crowd functions as access to market resources that allows ventures to gain and reinforce user base. Following this logic, then, the extent to which such crowd collaboration with external contributors can facilitate firms' potential for subsequent value capture hinges on the scope of users they can reach through the crowd. We argue that a diverse knowledge base of the firm expands the heterogeneity of external collaborators they can attract. As individual contributors also gravitate to the knowledge domains that they are familiar with (Foss et al., 2016), the knowledge breadth of the firm connects contributors for each of the domains within their knowledge repertoire. Those crowd collaborators, in turn, may impact on distinct user group within each of the domains, maximizing the scope of potential users the venture can reach. Moreover, the collaboration could become more effective as contributors and the firm already share similar technical background, language and mentality, which eases the communication and knowledge recombination process, further reinforcing the positive effect of crowd collaboration. In contrast, if firm's knowledge breath is limited, it may in fact repeatedly reach to the same crowd within a single knowledge domain, with all potential users subjecting to influence of the same crowd that repeatedly collaborate with the firm. In turn, the effect of crowd collaboration will be weakened as they can only reach to a niche market with limited market capacity.

H3: The venture's knowledge breadth in the open source communities strengthen the positive effect of collaborating with external contributors on the likelihood of receiving venture capital investment for an open source-based venture.

Figure 3.1 provides a summary of the hypotheses.

*** Insert Figure 3.1 here***

METHODS

Empirical context: open source software development communities on GitHub

We test the proposed hypotheses in the context of open source software development for several reasons. First, the open source software development industry is with the most active and developed open source innovation (e.g., Chesbrough, 2003; Foss et al., 2016; Von Krogh et al., 2003). As existing literature has demonstrated, firms also differ in their motivation of open source (Alexy & Reitzig, 2013; Hippel & Krogh, 2003), which may influence their propensity to participate in crowd collaboration after revealing their technologies and knowledge to the public. On the one hand, because this study focuses on the economic consequence of crowd collaboration post open source, the prevalence and heterogeneity of open source technologies and their sponsors allow us to empirically observe the variance in collaboration with the crowd contributors, which is the key to our theory and empirical analysis. On the other hand, the importance of open source technologies in this industry also makes the value creation a particularly salient and relevant issue in this context. A substantial amount of path-breaking technologies in this industry are released in the form of open source through permissive license rather than preoperatory innovation, many of which have profoundly influenced the entire economy and society such as cryptocurrency and artificial intelligence and deep learning data. The importance and volume of open source technologies hence make the economic value creation a critical to the development of firms and the overall industry.

Second, entrepreneurship and venture capital play a critical role in the development of the software technology industry. According to Crunchbase, 250 out of the 295 current unicorn ventures value at more than $1 Billion are associated with the software technology industries (including Internet, SAAS, Software, Artificial intelligence, Machine Learning, Cryptocurrency, Fintech, etc). According to CB Insights, more than 60% of the venture capital investment was made to ventures in the software technology industry, including internet, mobile applications, and other non-mobile software in 2017 and 2018. In several critical domains with emerging radical software technologies that are primarily open sourced, such as artificial intelligence, the overall investment increased 9 times from 2013 to 2017, from $1.15 Billion to overall $9.3 Billion (CB insights, 2018)[4]. The active entrepreneurship and venture capital investment also allow venture capital investment to be a representative proxy of economic value creation in this context. Figure 2 shows the intensity of venture capital investment made to ventures with open source activities by year from 2013 to 2017, the sample time of this study, which includes information on 10,742

[4] Source: https://www.cbinsights.com/reports/CB-Insights_MoneyTree-Q4-2018.pdf

rounds of investment. Table 3.1 provides some examples of open source-based new ventures in the sample, with their innovation, open source and investment information. Figure 3.2 shows the intensity of venture capital investment to open source based ventures during 2013-2017 the observation period.

*** Insert Table 3.1 and Figure 3.2 Here***

Data source and data collection

Our primary data source is GitHub.com, currently the world's largest host of computer source codes for open source software programs (Dabbish et al., 2012; Octoverse, 2018). GitHub started as a web-based cloud storage site for computer codes written a through distributed version control tool called Git. Initially developed by Linus Torvalds, who also created Linux, the world's most successful open source operating system that has been extensively studied in the literature (Henkel, 2006; Lee & Cole, 2003; Raymond, 2001) , Git was originally designed to function as a "content tracker" of each change made to the Linux kernel, so that developers can more efficiently track, compare and coordinate their development activities across different versions of Linux. In Git-based development process, the compete codebase is first mirrored to the developer's local computer, then each change (term as "commit") to the source code file of a technology has to be confirmed (term as "push") and submitted by the developer to the administrator or owner of the software program (term as "pull request") for review and approval, before it is officially incorporated to the default branch (Dabbish et al., 2012). Partly due to the profound influence and substantial user base of Linux, the technology quickly took off and become the dominant version control technology in software development (both open source and private), replacing other version control tools, such as source control management (which is proprietary) and Apache Subversion (SVN). Another similar web-based cloud storage service is sourceforge.net, a website of open source software that has been used as the empirical setting in technology management research (e.g., Foss et al., 2016; Wen et al., 2013). Sourceforge.net is equivalent to GitHub and has generated profound impact in the mid and late 2000s. While sourceforge.net hosts source codes with SVN based version control technologies, GitHub stores those created with Git (Wikipedia, 2018d). Figure 3.3 summarizes the software development process with Git technologies and the roles of pull requests and pushes in such process.

Insert Figure 3.3 here

Coincided with the dominance of Git and due to the importance of version control and discussion channels in open source technology development (Frederiksen & Rullani, 2016; Von Krogh et al., 2003), GitHub quickly gained popularity as the source code storage host for open source software technologies after initially launched in 2008, replacing sourceforge.net as the largest open source file host provider in the world in 2011 (Wired, 2016). Up to 2018, GitHub hosts the source codes of more than 96 million open source technologies, with more than 31 million active users and 2.1 million organizations participated in more than 200 million collaboration requests (Octoverse, 2018).

The basic unit of technology is referred to as "repositories" on GitHub. Each GitHub repository contains a set of program files (source codes) that constitute an open source software technology. At the same time, because GitHub provides the subscription function (users can receive the newsfeed) and discussion boards (for Q&A, suggestions, and communication) at the repository level, each repository essentially constitute the open source community for the focal technology, through which the sponsor (owner) of the technologies interact and collaborate with the crowd.

The open source repositories on GitHub are directly set under the user account of the initiating developer, which can be an individual or an organization. GitHub distinguishes the organization accounts from the individual accounts and provides information on whether the focal account belongs to an organization or individual to the public[5]. To better rule out the possibility that owners of the open source technology are not profit-seeking (individual hobbyist programmers) in a way that may render the economic value creation less relevant, in this study, we restrict our sample to the organization accounts and their affiliating repositories on GitHub. Among the organization accounts, the owners of open source technologies are composed predominantly of small ventures and organizations in the initial stage. However, it should be noted that more than 50% of the fortune 500 and 600 public companies also have public repositories on GitHub (Octoverse, 2018). The observation period of the study is from January 2013 to December 2017.GitHub became the absolute dominant open source host from 2013 (Wired, 2016), which allows us to observe the comprehensive set of open source collaboration ventures involved. We also include data only before 2018, to exclude the possible systematic change of the crowd's behavior caused by Microsoft acquisition on GitHub (Silver, 2018).

The primary data source is GitHub achieve, a publicly available database that compiles information on real-time GitHub activities. GitHub documents all user activities and repository information through its Application programming interface (API) service in the form of JSON (JavaScript Object Notation) files (Dabbish et al., 2012). GitHub allows public access and retrieval of data through such API, to obtain information on the updates and changes, discussions and collaborations for all public repositories. GitHub Archive downloads the JSON files on

[5] The website layouts of individual and organizational account are also different. An implicit assumption we made here is that the organizations setting up an organization account are with at least some intention of seeking profit and economic value.

development activities from GitHub API and compiles into data dumps on an hourly basis since 2012 (Grigorik, 2012). We supplement this data with another similar date source, GH Torrent, which further complies the real-time activity level JSON data into the relational database at the activity, project, individual and organizational level data (Gousios, 2013). At the same time, we used the GitHub API to directly retrieve the address of the official website, the critical information in mapping venture capital investment and ventures' open source activities, for all organizations with activities on GitHub prior to 2017 (a total 775,650 GitHub organizations accounts).

To measure those open source-based ventures' value creation, we rely on the venture capital investment data, obtained through the VentureXpert database on SDC Platinum. As previously discussed, we rely on the official website information, the only information available on both data sources, to map the GitHub organizations to investment data on VentureXpert.

Within our observation period, the mapping yields 10,742 rounds of venture capital investment mapped to 5,037 organizational accounts on GitHub. Together, the initial data contains longitudinal information on over the open source projects by 450,097 ventures, with 14,472,957 records of collaboration from 2013 to 2017, mapped to the investment data based on the address of company website.

Measures

Dependent variable

We measure the economic value of crowd collaboration through the variable *Venture capital investment*, which is the dollar amount of venture capital investment a venture has received up to a focal quarter to measure open source-based ventures. To account for the skewness of the distribution, we used log-transformed value of venture capital investment in regression estimations.

Independent variable

Pull requests. To measure ventures' collaboration with external contributors, we take advantage of a unique feature provided by GitHub to the open source technology, the "pull-request", a feature embedded in the original purpose of Git, the fundamental technology underlying GitHub. Essentially, the collaboration with external contributors for software development on GitHub is based on the Git version control technology. As a version control technology, Git mandates all the changes to the program source codes made by external members to be reviewed by the owner/sponsor of the software programs before they are actually incorporated as a function of the programs. Pull requests functions as the summary of pending changes initiated by the external developers (the crowd) who seek to participate in the development of a focal software program (repository). In order to send out pull requests, the external developers first need to create a copy the current version of the program to their own local directory (in their own GitHub account), on which they make changes to the codes. Pull requests are created once the external developers completed those changes and submit the codes back to the original repository for review. Owners or administers can evaluate the changes submitted through pull request before deciding on whether to accept or reject changes. In doing so, the pull requests allow both collaborators and owners to track the changes and keep the stability of the software programs by avoiding uncecessary or inappropriate changes made by external members. In the introduction of pull requests by GitHub, it explains "once a pull request is opened, you can discuss and review the potential changes with collaborators and add follow-up commits before your changes are merged into the base branch.[6]" Such way of collaboration, determined by the nature of the Git technology on GitHub allows us to observe measure crowd collaboration accurately. Figure 3.4A and Figure

[6] More details please refer to: https://help.github.com/en/articles/about-pull-requests

3.4B provide an example of collaboration process based on pull-requests, including the content of collaboration and the communication happened between a collaborator who initiated the pull request and the venture, Elastic. NV, an open source based venture with more than 400 open source technologies on GitHub that went IPO in October 2018, with price increased 94% to $ 72 per share on the trading(Solomon, 2018). The variable *Collaboration* is measured is the log of the number of pull requests initiated from external contributors within the quarter.

*** Insert Figure 3.4 A &B ***

Contingency variables

We measure venture's knowledge disclosure through the number of "*Pushes*", changes made by a venture to its source code repositories directly in the quarter. Similar to pull request, "push" is also an embedded function of updating programs files in the Git version control technology used by all software programs stored on GitHub. Different from "pull requests", a push is made by the owners of the repositories to upload content in their own local files (on their own terminal) to their public GitHub repository. Hence, the more pushes by the owner indicate the more development actives and source code the owner reveal to the public, which reflects the extent to which the owners' disclosure about their own knowledge. The variable *Pushes* is measured as the log of the number of the push commands send from the focal organization's GitHub account in the quarter.

We measure a venture's knowledge breaths based on the programming languages used in its software technologies. We focus on programming languages because the programming language is the fundamental knowledge component in our empirical contest, computer science related technologies. Each programming language represents a distinct set of logic and communication routines that can be extended and applied to a variety of functional domains (e.g., security, deep learning, web development, etc). The ventures and collaborators also need to invest substantial efforts in learning and experimentation in order to develop the source codes in certain functional domains using a focal language. Such characteristics of programming languages allow us to capture the knowledge breadth of the ventures in the context of computer science. More specifically, the measure is calculated as a Herfindahl index based on the programming languages used in all the repositories released by the venture up to the focal month. Models also control for venture experience and other knowledge related factors that may affect both venture capital investment and collaboration.

Control variables

Models also controlled for other factors that can affect venture capital investment to open source-based venture and can be correlated with crowed collaboration. First, the extent to which ventures' open source technologies can attract user interests may influence venture capitalists' evaluation on the technologies market potential and hence decision making of investment, while it can also affect the crowd's willingness to initiates collaboration through pull requests.

Hence, we controlled for user interest, measured as the number of *Watchers* who follow the updates on the ventures' technology through the "watching" function of GitHub. When "watching" the venture's repositories, the "watcher" will receive updates through the home page of their own GitHub account, a function similar to the "following" function on Twitter and other social media. The variable is calculated as the logarithm of the overall number of watchers that follows the repositories of a focal venture. Models controlled for the diffusion of ventures' source codes through *Forks*, number of times that the ventures' repertoires were duplicated to external developers' own account. In forks is a prerequisite of pull requests in the Git technology, it can affect the crowds' tendency of collaboration, at the same time, *Forks* are also commonly adopted by users to learn and explore the technology (Loyalka et al., 2019), which influences its market potential and hence economic value. Similarly, models account for the effect of the quality of ventures' technologies through the total number of *Issues*, the discussion board of repositories on GitHub most commonly used for reporting bugs and errors. All the variables take the logarithm forms to mitigate the possible influences to estimations caused by the skewness of the data.

In addition, ventures experience on GitHub can also affect collaboration and the extent to which they can create value based on open source technologies at the same time. To address this concern, models control for the *Number of libraries* published on the original platforms to the focal month, as well as *Tenure*, the number of days since the venture created the first repository on GitHub. Lastly, as our *Knowledge breadth* measure is based on the dispersion of programming languages, models consider the number of *Programming languages* used as the primary programming language, information automatically calculated and provided by GitHub API, in the ventures' repositories up to the focal month. To address the concerns of reverse causality, all the independent variables and control variables are lagged by 1 quarter to the dependent variables. In all estimations, the model also included venture fixed effect and month fixed effect. Table 3.2 summarizes the measurements of all variables used in the analysis.

***Insert Table 3.2 Here ***

Estimation strategy and sample selection: a difference-in-differences approach

In testing our theory, an empirical challenge is to account for the possibility that collaboration is not randomly assigned to ventures, which raises concerns with unobserved heterogeneity and reverse causality (Holland, 1986). That is, unobserved factors may exist that at the same time affect whether or not a venture collaborate with external contributors and also help explain the likelihood of receiving venture capital investment. Such issue is particularly salient given that the innovations disclosed by ventures may vary in their technological and market potential, which can influence both the crowd's willingness to collaborate with the venture, and the venture capital investment at the same time, leading to concerns of observed heterogeneity and omitted variable biases (Wooldridge, 2012). Although less likely, it might still be possible that the crowd became aware of venture capital investment to the ventures, and hence become more willing to collaborate in developing the open source technologies sponsored by those ventures, leading to the concerns of reverse causality.

To address such potential threats, we adopt a difference-in-differences approach in estimating the effect of crowd collaboration on the venture capital investment to the sponsoring ventures. More specifically, for each open source-based venture with crowd collaboration, we identify an otherwise similar open source-based venture without any crowd collaboration activity as the control venture.

Sample selection: propensity score matching

We rely on propensity score matching to identify the control group in the above estimations. For each treated venture that received crowd collaboration, we identify a counterfactual venture that is otherwise similar to the treated venture but without crowd collaboration through a probit estimation on the likelihood of receiving the treatment based on all independent and control variables. Table 3.3 reports the propensity score estimation by probit model. The control venture for a focal treated venture is identified as the venture using the 1 on 1 nearest neighbor matching, with the minimum difference in propensity scores predicted by the probit model in Table 3.3. The propensity score matching yields 64,590 out of 171,003 treated ventures at the quarter of treatment (receiving crowd collaboration for the first time), matched to 64,590 out of 279,094 control ventures with 2,781,745 quarter-observations.

*** Insert Table 3.3 here***

Table 4 compares the control and treated ventures before and after propensity score matching in terms of their venture capital investment, knowledge disclosure, and breadth and all control variables. Before the propensity score matching, the control and treated ventures exhibit drastic differences, with t-value over 100 in all criteria except for pre-treatment venture capital investment. The results indicate that the ventures receiving crowd collaboration also are significantly higher in knowledge disclosure and breadth, with technologies that gained higher user attention in terms of forks, issues, and watchers. The drastic differences verify the concerns of omitted variable biases, and hence further justify the necessity of matching, so that crowd collaboration can be regarded as a randomized treatment in deriving casual inferences. The three columns on the right of Table 3.4, show that, after the matching, such imbalances are greatly reduced, with no significant differences between the treated and control ventures in terms of *Issues* (as a reverse coded proxy for innovation quality) and *Tenure*. It should be noted that other criteria still demonstrate statistically significant differences, although the t-values are dropped by more than 90% compared with those pre-matching. Such statistical significance is partly due to the large sample size, which allows very small differences to be detected in the t-tests. Despite the significant t-values, the means very similar between the treated and control ventures in all those criteria after matching, with *Pushes* (knowledge disclosure), *Knowledge breadth*, *Forks* (as proxy for knowledge diffusion), and *Programming languages* only differ at 2 decimal level), which proves that the treated and counterfactual ventures are largely similar after matching. Figure 3.5 graphs the distribution of crowd collaboration likelihood based on the probit estimation. Consistent with Table 4, the treated group display much higher propensity of receiving crowd collaboration before matching. After matching, the distribution of receiving the treatment of crowd collaboration is almost identical for the treated and control ventures on GitHub. Such evidence increases our confidence in the effectiveness of matching, which allows us to regard crowd collaboration as randomly assigned treatment between the treated and control ventures in deriving the causal inference about its effects on venture capital investment.

*** Insert Table 3.4 and Figure 3.5 here***

Estimation models for difference-in-differences inferences (DiD)

We then estimate the effects of crowd collaboration on ventures' value creation by comparing the venture capital investment made to treated and control ventures within an eight-quarter (two-year) window before and after the treated ventures received the first collaboration request from the crowd (pull requests). In doing so, we implicitly assumptions that collaborating with the crowd does not necessarily co-exist with knowledge sharing and user communities of the open source technologies sponsored by the treated ventures. This assumption is in line both with our theoretical motivation about the challenges brought by collaborating with the crowd during knowledge sourcing mode, and with what we empirically observed from the GitHub data. As we will elaborate in greater detail, only a very small number of open source technologies are with crowd collaboration activities. By comparing those ventures with ventures that are otherwise similar but only without crowd collaboration, the analysis also focuses on the treatment effect of crowd collaboration, while minimizing the potential selection effect of crowd collaboration (i.e., the crowd are more willing to collaborate because of the quality and potential the technology).

The initial estimation model for a difference-in-differences approach takes the following form:

$$VC_{it} = \beta_a After_{it} + \beta_{ca} Collaboration \times After_{it} + BX_{it} + \vartheta V + \beta_i + \varepsilon_{it} \text{ (Equation 1)}$$

In equation 1 above, β_a captures the post treatment effect for all ventures. The coefficient β_{ca} captures the average treatment effect, that is, the difference in VC investment received pre- and pos- crowd collaboration between ventures subjected to treatment (crowd collaboration) and ventures in the control group (without crowd collaboration). B is a vector of coefficients on control variables $X_{i,t}$, ϑ represents the vector of coefficients on month dummies. It is to be noted that the control ventures in the models are open sourced ventures that share their technologies on GitHub, but without any crowd collaboration activities throughout the observation period.

Given the longitudinal nature of our data that allows us to observe the same ventures venture in both pre- and post-treatment periods, Equation 1 can then be simplified to

$$\Delta VC_i = \beta_a + \beta_{ca} Collaboration_i + \Delta BX_{it} + \Delta \varepsilon_i \text{ (Equation 2)}$$ [7]

Because crowd collaboration is measured by the number of pull requests, which is essentially a continuous treatment, Equation (2) is equivalent to:

$$\Delta VC_i = \gamma_a + \gamma_{pa} \Delta Pullrequest_i + \Delta \Gamma X_{it} + \Delta \varepsilon_i \text{ (Equation 3)}$$

Adding back the fixed effects in regression estimations leads to the following equation:

$$VC_{it} = \gamma_a After_{it} + \gamma_{pa} \Delta Pullrequest_{it} \times After_{it} + \Gamma X_{it} + \vartheta V + \beta_i + \varepsilon_{it} \text{ (Equation 4)}$$

[7] Equation 2 is derived from:

$$VC_{i1} = \beta_a After_{i1} + \beta_{ca} Collaboration \times After_{i1} + BX_{i1} + \vartheta Y + \beta_i + \varepsilon_{i1}$$
$$- VC_{i0} = \beta_a After_{i0} + \beta_{ca} Collaboration \times After_{i0} + BX_{i0} + \vartheta Y + \beta_i + \varepsilon_{i0}$$

In which 1 denotes post treatment periods and 0 denotes pre treatment periods

As the first instance of crowd collaboration happened at different time periods, we replace the dummy $After_{it}$with the actual quarter of the treatment (Woodrdige, 2010) and absorbed in to the fixed effects. Hence, our final estimation is based on the following equation:

$$VC_{it} = \varphi_p Post\_Pullrequest_{it} + \Gamma X_{it} + \eta Z + \beta_i + \varepsilon_{it} \text{ (Equation 5)}$$

In the equation, $Post\_Pullrequest_{it} = \Delta Pullrequest_{it} \times After_{it}$, Z represents time invariant fixed effects, including quarter fixed effects.

For testing the interactions as proposed in H2 and H3, we used the following model, in which the variable $Contigency_{it}$ refers to the contingent variables in H2 and H3. That is, *Knowledge Disclosure* (H2) and *Knowledge breadth* (H3).

$$VC_{it} = \varphi_p Post\_Pullrequest_{it} + \varphi_{pc} Post\_Pullrequest_{it} \times Contigency_{it} + \Gamma X_{it} + \eta Z + \beta_i + \varepsilon_{it} \text{ (Equation 6)}$$

In equation 6, ηZ also includes all the two-way interactions of the contingency variables with the quarter-fixed effects.

RESULTS

Table 3.5 presents the statistics and pairwise correlations for all variables. The average amount of venture capital investment received by the sample after log transformation is 0.09 (log unit $ Thousands), with a standard deviation of 0.9. Correspondingly, without log transformation, the mean of venture capital investment to the sample GitHub organization is $314.93 thousand, with a total of 1,989 venture capital investments made to 1153 organizations and an average amount of $23.45 million and highest firm value (as cumulative investment $ amount) at $5.3 billion. Despite the high variance, the summary statistics reveals the huge economic potential of open source technologies as reflected from venture capital investment. Similarly, the extent of crowd collaboration also exhibits considerable heterogeneity, with the number of pull request average at 1.18, a standard deviation of 0.6 and a maximum over 2,000 before log transformation. The high variance in crowd collaboration across open source technologies is in line with the theoretical motivation of this paper, showing that the extent to which ventures participate in crowdsourcing differ even after revealing their innovation as open source.

*** Insert Table 3.5 & 3.6 here ***

Table 3.6 presents the main regressions analyses based on fixed effects OLS estimations for the effect of crowd collaboration on the economic value creation potential of the open sourced based ventures. We used the Stata command "areg" to estimate the fixed effects, because this command takes into account the change in the degree of freedom after adding venture fixed effects (Wooldridge, 2012). Model 1 only considers the effect of control variables, including the fixed effect of ventures, quarter and the interaction of quarter with the independent and contingency variables. In Model 2, the independent variable that indicates the DiD effect, *Crowd collaboration* was added into the model. In Model 3 and Model 4 of Table 6, the contingency variables for testing Hypothesis 2 and Hypothesis 3, *Pushes* (for knowledge disclosure) and *Knowledge breadth* were added into the model separately. Model 5 is the final model that considers both the mean effects of crowd collaborations and the two constancies. In all models, the variables used in interaction terms are mean-centered and the VIFs are below the threshold of 10 and even the more stringent threshold of 5 (Kleinbaum et al., 1988), indicating that the models do not pose significant concerns of multi-collinearity despite the high R-squared due to the addition of complementor-fixed effects . The analysis is based on the final model with all fixed effects, the main effect and the contingency variables (Model 5 in Table 3.6).

Hypothesis 1 predicts that crowd collaboration increases the economic value open source-based ventures can potentially create, as reflected in its venture capital investment. Before turning into the results of regression analysis, we check the changes of venture capital investment before and after the treated ventures receive the first crowd collaboration request (pull requests) through graph analysis. Figure 3.6 shows the distribution of collaboration frequencies amount treated ventures. Figure 3.7 exhibits the mean of venture capital investment to treated and ventures up to the focal quarter before and after the first crowd collaboration of the treated open source-based ventures. While the treated ventures are with higher valuation even before the first crowd collaboration, the treated and control ventures demonstrated very similar trends in the growth of venture capital investment before the treatment of crowd collaboration. To be more specific, prior to the first crowd collaboration of the treated ventures, both control and treated ventures experienced slow and fluctuating increase in venture capital investment, providing evidence to the equal-trends assumption required by the difference-in-differences design (Angrist & Pischke, 2008). On the other hand, in quarters after the treated ventures started to receive crowd collaboration, investment to the treated ventures grow more steadily, while the investment to the control ventures largely maintained the similar pattern in the pre-treatment period. As a result, the gap in the growth of venture capital investment between the treated and control open source- based ventures widened after the treated ventures' first instance of crowd collaboration. Thus, the graph provides preliminary support to Hypothesis 1.

*** Insert Figure 3.6 & 3.7 here***

Now turning to the regression analysis, consistent with the Hypothesis 1 and the graph analysis, the coefficient of *Crowd collaboration* in Model 5 of Table 3.6 is significant ($\beta = 0.076$, $p < 0.001$). As both the dependent variable, thousand-dollar amount venture capital investment, and the independent variable of *Crowd collaboration* takes the form of log transformation, the coefficient indicates even if the number of *Crow collaboration* increases by 1 from the mean (which is almost 100% increase from the mean of 1.18), the economic value potential of the ventures reflected through venture capital investment will increase around 8%. Such results provide strong support to Hypothesis 1.

Hypothesis 2 predicts the number of *Pushes* that ventures made to disclose their own knowledge will attenuate the positive effect of crowd collaboration on venture capital investment. In line with this prediction, the coefficient of *Crowd collaboration X Pushes* is negatively significant ($\beta = -0.013$, $p < 0.001$), showing that the economic value potential of Crowd collaboration can be offset by the opportunity cost of forfeiting the proprietary rights of ventures' own knowledge. Hence, Hypothesis 2 is supported.

Hypothesis 3 argues that the positive effect of crowd collaboration on venture capital investment to the open source-based ventures is stronger the ventures with broader knowledge. Model 5 in Table 3.6 reports a positively significant coefficient of *Crowd collaboration X Knowledge breadth* ($\beta = 0.055$, $p < 0.001$). It shows that ventures whose repositories contain a diverse set of programming languages received even more venture capital investment post crowd collaboration, supporting the argument of Hypothesis 3. In summary, the regression analyses show full supported to our proposed hypotheses.

Robustness checks

We further conducted additional analysis to test the robustness of the results. We first investigated whether the observed effect of crowd collocation is sensitive to model specification. To do so, we tested the model excluding the fixed effects and using alternative linear models. In Table 3.7, Model 1 reports the final model (Model 5 of Table 6). Model 2 and Model 3 report estimations that exclude all fixed effects and excludes only the interactions of the fixed effect dummies with independent/contingency variables respectively. Model 4 of Table 3.7 presents the analysis of the model without adjusting the degree of freedom associated with the fixed effects dummies using Stata comment “xtreg”. The results remain fully robust to those alternative specifications of models and estimations strategies.

*** Insert Table 3.7 here***

Second, we tested the sensitivity of the effects of *Crowd collaboration* to the 8-quarter time window specification pre- and post-treatment in our main model using alternative specifications of observation windows. More specifically, we estimate the using alternative time windows as reported in Model 5-7 of Table 3.7, with the same set of variables in Model 5 of Table 6 (final model). Model 5 is based on a six-quarter time window, and Model 6 is based on a period of ten-quarter time window before and after the treatment of *Crowd collaboration*. To address the concern that the results may be influenced by the proportion of ventures exist less than six quarters before receiving the treatment of *Crowd collaboration*, we also performed analysis using an unbalanced time window from 2 quarters pre to 8 quarters post the first instance of Crowd collaboration in Model 7 of Table 3.7. Overall, those results are highly consistent with the main analyses, providing additional support to the robustness of our conclusion.

Lastly, to rule out the possibility moderate correlation between the contingency variable *Pushes* and independent variable *Crowd collaboration* ($r = 0.43$ in Table 5) may cause biases in estimations (Kalnins, 2018), we tested the model using the alternative calculation of *Pushes*. Note that the VIFs in all our models are below the conservative threshold of 5, indicating low risks of biases in coefficients due to multicollinearity. In Model 8 and Model 9 of Table 3.7, we further replaced the log-transformed number of *Pushes* (which better approximates normality and reduces skewness of the data) with the original form of the number of *Pushes* in measuring the knowledge disclosure. The results are consistent with our main analyses, further mitigating the concerns of potential biases due to correlation and multicollinearity.

Testing the underlying mechanisms: crowd collaboration and market resources

One limitation of our empirical analysis is that we are not able to directly test the effect of crowd collaboration on the market share of ventures' technologies, which is the main underlying mechanism proposed in the theory development. Due to the open source nature of those technologies, it is impossible to fully observe the market share or user base for the majority of the ventures in our sample. However, to further explore whether the observed effect of crowd collaboration on the venture' valuation is indeed driven by our proposed mechanism related to expanded market resources, we tested how crowd collaboration impacts on the total number of *Forks* and *Watchers* of ventures' repositories in the subsequent quarter.

Although they do not represent the entire set of market resources, the two measures, to a large extent, can effectively proxy the market resources that are critically relevant to the two underlying mechanisms of crowd collaboration in the theory development, regarding increasing the stickiness of lead users and the second order effect of general users. First, *Forks* can reflect the wiliness to learn ventures' technologies by users with adequate knowledge, a critical step in attracting and retaining the lead users. Once forking a repository, the users essentially copy the entire source code of the technology to their local directory, an action that would not be necessary if the intention of the user is only to use the technology unless they want to explore the detailed programming content. Hence, *Forks* can proximate the potential scope of users that can become entrenched to the ventures' technologies due to the familiarity of knowledge and routines, as argued in the theory development. *Watchers*, on the other hand, constitute a broader scope of users who show general interests in the focal technologies. Hence, it can proxy the extent to which ventures are able to attract general users as their market resources as a result of crowd collaboration, another main mechanism proposed in the theory development.

*** Insert Table 3.8 here***

Table 8 presents the results of the Differences-in-Differences model using the log or *Forks* and *Watchers* as the dependent variable. Model 1 and model 2 are estimations with *Forks* as the dependent variable, and Model 3 and Model 4 are estimations with *Watchers* as the dependent variable. Model 1 and Model 3 consider the main effects of crowd collaboration, and Model 2 and Models report the full model with contingency effects. In all models, the specification of control variables is the same as Model 2 and Model 5 in Table 3.6 (main results table). Same as our main estimations, all the independent variables and control variables (excluding fixed effects dummies) are legged by one-quarter to the dependent variable to avoid reverse causality. As show in the Table, crowd collaboration has a significantly positive influence on both *Forks* and *Watchers* ($\beta = 0.195$, $p < 0.001$ in Model 2 for *Forks*, and $\beta = 0.199$, $p < 0.001$ in Model 4 for *Watchers*), providing further evidence to our underlying mechanisms. Similarly, the contingency variables, developed based on the two underlying mechanisms, also display statistically significant patterns on *Forks* and *Watchers* that are consistent with our theory. In line with its impact on venture capital investment to the ventures, knowledge disclosure through *Pushes* weakens the positive impact of *Crowd collaboration*, with significantly negative coefficients of *Crowd collaboration X Pushes Watchers* ($\beta = - 0.032$, $p < 0.001$ in Model 2 for *Forks*, and $\beta = - 0.013$, $p < 0.001$ in Model 4 for *Watchers*). Similarly, *Knowledge breadth* reinforces the positive effect of *Crowd collaboration* on gaining those market resources, with significantly negative coefficients of *Crowd collaboration X Pushes Watchers* ($\beta = 0.077$, $p < 0.001$ in Model 2 for *Forks*, and $\beta = 0.053$, $p < 0.001$ in Model 4 for *Watchers*). In summary, those supplemental analyses further provide evidence to the validity of the mechanisms related to market resources underlying how crowd collaboration can create economic value for open source-based ventures.

Alterative explanations

In the last set of supplementary analyses, we seek to rule out the alternative explanation that drives the observed effect of *Crowd collaboration* on venture capital investment. While our theory argues that the crowd collaboration allows ventures to create value by accumulating rare and valuable market and user resources, a salient alternative explanation is that the crowd could still be valuable as upstream knowledge resources that allow firms to increase the quality of their open source technologies, increasing the potential of value capture. To rule out this alternative explanation and further verify the mechanisms we propose, we took advantage of the fact that our data actually separates the initiation and completion of crowd collaboration. While our measure in the main analysis focuses on the number of pull requests opened (that is, the number of collaborations initiated by the crowd), the data also provides information whether the venture decided to "merge" the pull results (that is, incorporating the source codes contributed by the crowd to the main program disclosed in their repositories). While our argument about the market resources should happen at the initiation (opening) of a pull request, the pull requests have to be merged into the repository in order for the crowd contributions to become actual knowledge sources for the venture. In Model 1 of Table 3.9, the coefficient of *Collaboration accepted* is positively significant ($\beta = 0.017$, $p < 0.01$), which is consistent with the original measure of crowd collaboration that takes account into all pull request initiated. However, as shown in Model 2 and Model 3 of Table 9, when considering all pull request opened, the effect of *Collaboration accepted* disappeared (β = - 0.003, p = *n.s.*). In other words, whether the ventures decided to utilize the knowledge from the crowd does not seem to matter after considering the crowds' willingness to collaborate, indicating that the value created by the crowd does not reside in its actual knowledge contributions to ventures' open source innovation. Hence, such evidence allows us to rule out the alternative explanation of the crowd as valuable knowledge resources, further strengthening the validity of our proposed theoretical mechanism (crowd as valuable downstream market resources).

*** insert Table 3.9 here***

DISCUSSION

This paper explores how the collaboration in open source with crowd contributors affects venture capital investment to open source based new ventures. Our inquiry is driven by the puzzle that the open source knowledge obtained through collaboration is publicly available, which runs counter to a central tenet in strategy research that private knowledge is key to the superior innovation that can maximize value capture and hence attract venture capital investment (e.g.,Barney, 1986; Peteraf, 1993).

In theorizing the impact of external contributors on venture's acquisition of venture capital investment, we highlight the role of the crowd in brokering the market resources for open sourced based ventures. More specifically, we argue the crowd collaboration first allows the ventures to reinforce the crowd as their own user base as their familiarity of the technologies and the development routines increase. Second, we argue that the crowd as lead users also plays a critical role in technology diffusion to ordinary users due to their importance in creating direct network effects among users. Therefore, collaborating with the crowd allows the ventures to retain resources critical to the competition of their innovation market share, which is essential to the value capture of ventures' open source technologies. Such role of the crowd in establishing market resources ultimately translates their free knowledge contribution into financial value by attracting venture capital investment.

Limitations

One major assumption we made in identifying the samples is that the organizations on GitHub seek economic profit. To mitigate this concern, we restricted our sample only to repositories under organizational accounts, to exclude the possible confounding influence on our estimations due to the existence of individual hobbyists (Foss et al., 2016; O'Mahony, 2003). Due to the large sample size, however, we acknowledge that so far we are unable to distinctly to determining the extent to which the organization accounts are seeking economic value and more specifically venture capital investment. Yet, it should be noted that such assumption of profit-seeking has become increasingly valid within our empirical context GitHub, even at the individual level. To validate this assumption, we conducted interviews with entrepreneurs and individual developers, who confirmed that the GitHub account has been increasingly unutilized to vet the capability of developers during job hunting. Similarly, the exponential growth of organization accounts (Octoverse, 2018) is largely rooted in the needs for ventures to showcase and gain the trust of the customers in terms of the quality and stability of their technology (Fosfuri et al., 2008). All those motivations that drive the activities are closely associated with profit-seeking of individuals and organizations.

Another potential limitation of this study is that we are unable to observe other activities of the focal ventures beyond open source communities. It is possible that venture capitalists invest in the ventures with open source repositories on GitHub, but not because of their open source technologies. For example, the ventures may have developed proprietary technologies at the same time, which allow them to directly profit from their innovation in a way that drives venture capital investment. However, as the current literature has demonstrated, open source software technologies were initially developed against proprietary innovation (Von Krogh & Von Hippel, 2003). Such institutionalized ideology makes it unlikely for open source based ventures to patent their innovation concurrently, especially given the increasing popularity of permissive open source license, such as MIT license, that waives all restrictions under which the usages of a technology is regarded (as opposed to the more prevalent GPL license in the 1990s and early 2000s). To further probe for this possibility, we randomly sampled ventures with GitHub responsories to examine whether those ventures have patenting activities in USPTO database. However, we did not find any patenting activities for the selected ventures, which helps mitigate the concern that venture capital investment to the sample venture might be associated with innovation activities beyond open source and GitHub.

Managerial implication

This study bears important managerial implications to entrepreneurship in terms of strategies to manage open source communities. The study demonstrates that, to maximize the value capture potential of open source technologies, ventures may need to go beyond simply showcasing ventures' technology and capability through open source. Rather, despite that the knowledge provided by the crowd contributors can be amorphous and difficult to absorb, it is worth maintaining the open source community for their technologies and involving the crowd in the development process, because of the value of the crowd in establishing for downstream market resources. Doing so would require venture to devote constant effort to lead, cultivate routines for coordination, and actively communicate with the open source communities, rather than merely disclosing the source codes.

The contingencies in Hypothesis 2 and Hypothesis 3 also bare implications for ventures in managing the knowledge in open source community platforms. More specifically, the results reveal that disclosing ventures' own knowledge attenuate the positive effect of collaboration, while a broad scope of disclosed knowledge accentuates such positive effect. Such results highlight a challenge ventures may face in open source innovation – to strike a balance between minimizing the amount and maximizing the scope of knowledge disclosed – in order to fully exploit the value creation potential of their open source technologies.

At the same time, our study reveals an interesting caveat for managing crowd collaboration on open source – ventures may not need to incorporate the actual knowledge from the crowd. In the supplementary analysis, we show that the merging pull requests from the crowd does not further increase the venture capital investment, once we account for how many pull requests are initiated by the crowd. This finding may have implications for open source-based ventures in terms of how to manage their internal knowledge with external knowledge from the crowd in the process of innovation.

Contribution and future research

This study first advances the understanding of crowd collaboration in open source. One contribution of this study is the explicit emphasis of the crowd as downstream resources for market competition. While the existing literature on both crowdsourcing and open source communities both regards the crowd as resources for knowledge inputs (e.g., Afuah & Tucci, 2012; Bayus, 2013; Boudreau & Jeppesen, 2014), it also notices that the crowd may not consistent an efficient access of knowledge for organizations due to non-contractual nature of such collaboration as well as organizations' own limitation in search and absorptive capacity (Piezunka & Dahlander, 2015; Von Krogh et al., 2003). Such limitation brings the necessity of crowd collaboration into question. In this study, we seek to this tension by focusing on the value of the crowd in the downstream market competition. More specifically, our theoretical development shifts the focus away from upstream knowledge sourcing and highlights the positive role of the crowd during technology diffusion and competition. Our theory and empirical analyses reveal that it is such process that the crowd facilitates a higher possibility of economic value creation as reflected in the venture capital investment made to open source-based ventures.

In addition, our study also highlights the heterogeneity of crowd collaboration among open source innovation, which, surprisingly is rarely addressed in the current literature. Indeed, most of the research on open source tend to regard crowd collaboration as inherent to open source technologies, with the assumption that the disclosure of knowledge can naturally attract users and external developers to discuss and make contribution to the technologies (Bianchi, Kang, & Stewart, 2012; Oh & Jeon, 2007). This dissertation' theoretical tension, emphasizes the possibility that because the knowledge from the crowd lacks consistencies and may be too distant to absorb, ventures may intention stay away from crowd collaboration for knowledge creation, even when they are willing to open source their own innovation for other benefits. From this perspective, it highlights crowd collaboration as a strategic choice of ventures. At the same time, our empirical evidence further provide evidence that the crowd collaboration could be a process that is distinct from open source as knowledge sharing, given the surprisingly low average number of crowd collaboration in our sample. Such empirical observation opens up many possibilities for future research to further explore factors driving such heterogeneity of crowd collocation in open source.

Second, this study deepens the understanding of value creation for open source innovation, especially with regard to the role of community and the crowd. How organizations can profit from open source innovation has been at the center of discussion in the open source literature (Bonaccorsi & Rossi, 2003; Fosfuri et al., 2008). Most of the existing research focuses on two distinct approaches of value creation from open source innovation. The first is the business model innovation based on open source technologies (e.g., Bonaccorsi et al., 2006; Casadesus - Masanell & Zhu, 2013), including providing charged services, coordinating technology standard and attracting priced complementors, etc. The second approach focuses on the dynamics between open source and proprietary innovation and highlights the value of open source in organizations' search of patentable technologies (e.g., Alexy & Reitzig, 2013; Hippel & Krogh, 2003). In those discussions, how the open source communities, a critical force that drives the development of open source technologies (Kogut & Metiu, 2001; Lerner & Tirole, 2005a), can play a role in the value creation is rarely addressed. Rather, open source communities are often regarded as independent out of the value creation process, as platforms in which crowds use and contribute to the technologies for free, and with strong ideologies against commercialization (Von Krogh & Von Hippel, 2003, 2006). This study seeks to resolve those conflicting views about the crowd, communities and the commercial value of open source innovation. In particular, the theory focuses explicitly on the role of the crowd collaboration and open source communities in educating and lock-in users, especially lead users, in a way that benefit the ventures in the downstream market competition that directly influences the economic value creation potential.

In addition, our focus on venture capital also deepens the understanding of how the value creation is evaluated for open source innovation. As the first study explicitly focusing on venture capital investment to open source new ventures, we shed lights on about how those ventures, as well as the value of their technologies, are assessed by venture capitalists. Moreover, we also connected the development process with the evaluation of a venture's economic value, when the value capture cannot be directly realized by innovation. While our study focuses exclusively on open source-based new ventures, future research may explore how the choice of open source, as opposed to commercialized close innovation, would affect venture capital investment, and more importantly, whether venture capital's preference of open versus closed innovation shifts, as open source technologies become increasingly influential over time.

Essentially, crowd collaboration can be regarded as a special form of inter-organization collaboration in the unique environment of open source communities (Afhua). From this perspective, this study also advances the understanding of the value creation processes of inter-firm collaboration. First, the study focuses on the financial impact of collaboration on new ventures. While an extended literature on alliances has investigated the stock market reactions to alliances announcement to gauge the value created through such inter-firm collaboration (e.g., Anand & Khanna, 2000; Doz, 1996), it exclusively focuses on such process for established firms, who inherently faces less uncertainty in both market demand and innovation during collaboration. The research on the interfirm collaboration of nascent firms, on the other hand, only noticed the role of venture capital in facilitating new ventures' subsequent collaboration (e.g., Colombo & Grilli, 2010; Davila, Foster, & Gupta, 2003; De Clercq & Sapienza, 2006). How venture capital evaluating new ventures' existing collaborative relations and their financial value in the first place, on the other hand, is rarely explored in the existing literature. By discussing the impact of collaboration in open source, this study sheds lights on the potentially critical role of a venture's existing collaborative network in creating financial value in a relatively conservative setting. Even when the knowledge resources from collaborators is not rare and valuable alone, as they provided from free without proprietary rights, we show that new venture can still financially benefit from such collaboration as they help new venture to better navigate innovation for market demand and make ventures' innovation non-substitutable by increasing the potential of a technology ecosystems surrounding the ventures' technology.

Secondly, this study depicts a unique value creation process from open source collaboration. Unlikely alliances under well-specified contracts and intellectual property (e.g., Gulati, 1995; Williamson, 1979; Williamson Oliver, 1985), open source is not bounded in terms of the scope of collaboration, nor by explicit ex-ante contracts. Most importantly, such collaboration allows free flow of knowledge from the new ventures to external collaborations as the underlying knowledge is fully disclosed to the public. Existing literature has yet explored how those features unique to open source would cause the evaluation process of its financial implication to differ. The theorization in this study is the first attempt to explore the unique mechanism of value creation underlying collaborative open source innovation. Rather than accessing and adapting to collaborator' knowledge as depicted in inter-firm proprietary innovation research, this study highlights that, in open source, the financial implication of collaborating with the crowd also is rooted in its influence on the downstream market competition, when knowledge resources are free and easily accessible.

Another potential for future research derived from this study is to explore other boundary conditions in the effect of open source collaboration on venture capital investment. The discussion of contingencies focuses on characteristics of potential collaborators in the contributor pool. Other external factors may also influence the extent to which venture capitalists value open source collaboration. One potentially fruitful consideration is the characteristics of the technological domain of the ventures. It is possible that the effect of collaboration on venture's financial potential is particularly strong at an early development stage of a technology, where the market demand is highly uncertain and the potential for building surrounding technology ecosystems are particularly high. Under such situation, the two proposed mechanisms underlying collaboration can play a more prominent role in determining the ventures' financial competitive advantage. Similarly, the effect of collaboration can also be particularly high in open source based technological field lead by incumbent firms. Without collaboration, it is more difficult for new ventures to compete with incumbents who possess complementary assets and are better able to develop complementary technologies.

The last remaining question is the comparison between open source and proprietary collaboration in creating financial value for new ventures. it is worth inquiring in subsequent studies, would open source collaboration more efficient than collaboration in closed forms for ventures in innovation? Relatedly, how would the financial value of the innovation and the venture differ because of the proprietary/open source nature of collaboration? While this study focuses exclusively on open source-based ventures and the implication of collaboration in the open source environment, it shed lights on those directions future research can pursue.

CONCLUSION

This dissertation explores how open source innovation, the practice in which firms distribute their technologies and share the underlying knowledge to the public without claiming proprietary rights, creates economic value for entrepreneurial firms. The motivation of this dissertation originates from the contrast between the increasing prevalence of open source technologies among profit-seeking ventures in practice and the fundamental emphasis on intellectually property rights in the existing literature on profiting from technology and innovation. The particular focus on entrepreneurial and new ventures is rooted in their importance as the driving forces in open source innovation(Wen et al., 2015), and the resource constraints they face makes the implication of open source particularly relevant. More specifically, on the one hand, open source allows new ventures to gain access to a variety of critical resources that can be otherwise difficult to obtain (including knowledge inputs, access to users, etc), on the other hand, by forfeiting the proprietary rights of their technologies, ventures may give up the most critical advantage that enables economic value creation from their innovation. Seeking to resolve this tension, the overarching research question of this dissertation is: why and how open source innovation create value for new ventures?

In investigating this question, this dissertation conceptualizes the development and competition among open source technologies as based on multi-sided open platforms, on which the shared non-proprietary knowledge constitutes a basis to attract additional knowledge inputs (the supply side of a platform) and users that seek to utilize and exploit the knowledge (the demand side of a platform). Accordingly, the review on three streams of relevant literature (1) open source innovation (2) platforms and two-sided markets (3) technology ecosystems identifies several important gaps in the existing studies. First, the literature on open source innovation tends regard open source technologies as distinct from commercial technologies due to their non-profit seeking nature, with unique dynamics in ideology, communication, governance and knowledge creation within the platform-based communities. Even though a few studies have explored the possibility of profiting from open source innovation, they tend to focus on the subsequent business model innovation without discussing the role of the technology community platforms that created and sustain the development of ventures' innovation. Second, the literature on two-sided markets and platforms tend to focus on the strategies and implications for platform owners priced technologies and products, without much discussion on how other actors, particularly complementors, can compete on two-sided platforms. The emerging literature on technology ecosystems, on the other hand, focuses largely on within-ecosystems dynamics with the underlying assumption about the critical importance of intellectual property rights. Those gaps in the existing literature further highlight the importance of the research question of this dissertation – how ventures, taking different roles as owners or complementors of open source technology platforms, can gain economic value of their open source technologies when multiple open source technologies and platforms are competing with each other?

Seeking to address those tensions in the current literature relevant to the overarching research question of this dissertation, the two empirical essays in this dissertation focus on how the dynamics and strategies unique to the non-proprietary community platforms shape the value creation of ventures' open source technologies. The first essay focuses on the expansion strategy of ventures as complementors to open source platforms. More specifically, it investigates the effect of multihoming, complementors' strategies to provide a similar set of complementary technologies to multiple platforms, on the complementors' growth on the user base, a pre-requisite of value creation on its original platform. The theoretical development of the essay highlights the possibility that multihoming transfers the platform network effects across multiple platforms, as it allows user to benefit more extended scope of communication (the transfer of direct network effects among platforms), and lower the overall cost of learning in adopting the complementors' technologies (the transfer of indirect network effects among platforms). As a result, both mechanisms allow complementors to reinforce their user base on the original platform after multihoming. It should be noted that although the essay focuses on the open source platforms without price mechanisms, the conceptualization of inter-platform transfer of network effects can also be generalized to other platform-based innovation and competition contexts where the price mechanism may exist.

The second essay focuses more exclusively on the unique dynamics in open source community platforms and directly links strategies ventures can leverage on those platforms with economic value creation. The essay extends the conceptualization of open source platforms from two-sided platforms, where transactions among complementors and users are mediated by a same common technological infrastructure provided by the owner of the platform, to multi-sided markets, where exist another type of critical actors, the crowd contributors. More specifically, the essay investigates how collaborating with the crowd contributions allows ventures to create economic value from open source innovation. The research question originates from the puzzle in the existing literature that the crowd as free resources for knowledge inputs may not create competitive advantage for ventures, while it has been shown that ventures also have difficulties assimilating the scattered knowledge from the crowd. In contrast with such notion of the crowd as free knowledge puts, the essay highlights the impact of crowd competition downstream, during market competition. More specifically, it argues that collaborating with the crowd allows the ventures to gain access to market resources critical to value creation, as the collaboration process develops path-dependencies and inertial for lead users that increases their stickiness to ventures' technologies. In presence of direct network effects on those open source platforms, those lead users, who are locked into the ventures' technologies through collaboration, play a particularly prominent role in user adoption and technology diffusion, which in turn allows ventures to gain access to a broader scope of ordinary users through crowd collaboration.

The empirical analyses of the two essays are based on a set of unprecedented rich and detailed data of the open source software industry from GitHub.com, the largest open source software storage host in the world with over 96 million technologies, 2 million organizations and 200 million collaborations among 31 million software developed. The data is updated on a daily basis, with detailed records at the development activities level, with over 5 Terabytes of information. Leveraging the advanced cloud computation and big data analytics technics, with research design that seek to derive causality, the empirical findings provide full supports to the theoretical hypotheses of the two essays. Together, the theory and findings of this dissertation seek to advance the understanding of value creation from such distributed and platform-based innovation process of open source in the digital era.

List of Tables

Table 1.1. Summary of literature on open source innovation

Dimensions	Central question	Major factors	Example studies	Limitations and Unaddressed questions
Individual motivation	What external collaborators are willing to contribute their knowledge to open source communities for free?	Intrinsic motivation (Ideology, need, interest)	Hertel, Niedner, & Herrmann, 2003; Bagozzi & Dholakia, 2006; Krishnamurthy, Ou, & Tripathi, 2014; Von Hippel and Krogh, 2003	• Focuses on signal open source technology/community • Why contributors choose some communities to contribute over others? How the heterogeneous characteristics of contributor motivation affect governance choice and effectiveness? • The specific impact of contributors on innovation • The evolution of contributor motivation over time
		Extrinsic motivation (career visibility, status, reciprocity)	West, 2003 ; Alexy, Henkel *et al.*, 2013 ; Krishnamurthy, Ou *et al.*, 2014	
		Environments characteristics (project size, governance, open source license)	Shah, 2006; Oh & Jeon, 2007; Belenzon & Schankerman, 2015; Foss, Frederiksen, & Rullani, 2016	
Innovation process	How innovations are created in open source?	Coordination (motivating, learning and adaption with the crowd's knowledge)	Lee & Cole, 2003; Dahlander & Magnusson, 2005; O'Mahony & Ferraro, 2007	• Focuses on signal open source technology/community • No connections about the four elements in the innovation process • Competitive and technological consequences of control through open source license
		Communication (content and method)	Lee & Cole, 2003; Bagozzi & Dholakia, 2006; Dahlander & Frederiksen, 2012	
		Control (knowledge disclosure, decision rights, open source license)	Kogut & Metiu, 2001; Henkel, 2006; Kuk, 2006; Alexy, George, & Salter, 2013	
		System design (modularizablilty; technology complexity)	MacCormack, Rusnak et al., 2006; Alexy, George et al., 2013; Baldwin & Clark (2006)	
Competitive dynamics	Can open source compete with proprietary innovation? When open source can be superior governance mode of knowledge sourcing	Antecedence/ decision of open source by firms	Afuah & Tucci, 2013; Almirall & Casadesus-Masanell, 2010; Wen, Ceccagnoli et al., 2015	• The sequence of private investment and open source • Mixed conclusion about the performance implications of open source that contracts reality • How open source technologies compete with each other.
		Performance of open vs. closed innovation	Economides & Katsamakas, 2006; Wen, Forman et al., 2013Piezunka and Dahlander (2015)	
		Value appropriation	Bonaccorsi, Giannangeli, & Rossi, 2006; Henkel, 2006; Alexy & Reitzig, 2013	

Table 1.2. Comparison of research on open source innovation communities, platforms and ecosystems

Technology systems	***Focus of analysis***	***Unique mechanisms proposed***	***Price assumption***	***Technological interdependency assumption***
Open source communities	Technology, within system	Innovation without proprietary rights	No price signal	High interdependency, only within system
Platforms	Market, within and between system	Network externality	Price as the fundamental determinant in market behavior	No technological interdependency
Ecosystems	Technology, within system and supply side	Technological interdependency/ system structure	Proprietary and priced technologies	High interdependency, mostly within system

Table 2.1. Examples of libraries on platforms

Library name	***Core technology***	***Publishing package manager platform***	***Description***	***User base***	***Own dependencies***
rake	Ruby	Rubygems	Software task management and build automation tool.	548K	0
rails	Ruby	Rubygems	Server-side web application framework	423K	0
express	JavaScript	NPM	Web application framework	546K	30
coffee-script-source	CoffeeScript	JavaScript	Secondary simplified language that compiles into JavaScript.	359K	0
mocha	JavaScript	NPM	JavaScript test framework featuring browser support, asynchronous testing, test coverage reports, etc.	331k	10
requests	Python	Pypi	Python library for HTTP requests	91.2K	0
numpy	C	Pypi	Array processing for numbers, strings, records, and objects	41.6K	0
Keras	Python	Pypi	Deep Learning for humans	3.74K	0

Description summarized based on information from GitHub and Wikipedia

Table 2.2. Summary of measures in Chapter II

CONSTRUCTS	***VARIABLE***	***MEASURE***
Dependent variable		
User adoption	Dependencies	Number of users' source code repositories containing a focal library as dependency
Independent variables		
Treated venture	Cross-platform complementor	Binary variable set to one if a complementor published its libraries on two package manager platforms during the period of observations
Time of expansion	Multihoming	Binary variable set to one if the complementor has published libraries on an additional platform other than its original platform
User base	Watchers	Number of libraries that can be jointly used with the focal library
Complementor's reliance on platform knowledge	Own dependencies	The number of dependencies specified by the libraries' own manifest file
Relative platform completive advantage	Relative platform size	Ration of number of users of the original platform to the newly entered platform
Control variables		
External collaboration	Pull requests	The number of external contribution submissions to a libraries' source file
Innovation quality	Issues	The number of bugs and errors in the 'issue' section of the library's GitHub page
Ventures' development efforts	Commits	Number of times the complementor modifies the source files of the library
Diffusion	Forks	The number of developers copied the source file of the library to their own repositories
Ventures' technological capabilities	Number of libraries	Number of libraries published on the original platforms
Venture experience	Tenure	The number of days since the complementor created the first library that was updated to the original package manager platform

Table 2.3. Probit estimation on the likelihood of complementor multihoming

DV	Likelihood of multihoming
User adoption	0.0065***
	(0.0010)
Watchers	0.0230***
	(0.0054)
Own dependencies	0.0083***
	(0.0023)
Pull requests	-0.0027
	(0.0455)
Issues	0.0187*
	(0.0096)
Forks	0.1052***
	(0.0034)
Commits	0.0044
	(0.0041)
Number of libraries	0.0015***
	(0.0002)
Create time	-0.0004***
	(0.0000)
Tenure	-0.0011***
	(0.0000)
Constant	5.6768***
	(0.1067)
Observations	6,151,672
Platform dummies	Yes

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

Table 2.4. *t*-Tests of user adoption and library features before and after matching

		Unmatched sample			Matched sample		
VARIABLES	All	Control	Treated	t-value	Control	Treated	t-value
User adoption	1.167	1.168	0.996	13.574	1.595	1.649	-1.970
		(0.001)	(0.011)		(0.020)	(0.019)	
Watchers	0.077	0.077	0.110	-15.900	0.153	0.158	-0.775
		(0.000)	(0.003)		(0.004)	(0.004)	
Own dependencies	0.927	0.928	0.831	14.992	1.183	1.230	-3.087
		(0.001)	(0.007)		(0.011)	(0.011)	
Pull request	0.000	0.000	0.002	-9.147	0.002	0.002	0.062
		(0.000)	(0.000)		(0.000)	(0.000)	
Issues	0.024	0.023	0.043	-18.433	0.053	0.056	-0.818
		(0.000)	(0.000)		(0.002)	(0.002)	
Forks	0.128	0.127	0.167	-15.400	0.267	0.247	2.737
		(0.000)	(0.003)		(0.006)	(0.005)	
Commits	0.081	0.081	0.132	-20.510	0.174	0.190	-2.129
		(0.000)	(0.003)		(0.005)	(0.005)	
Number of libraries	3.632	3.635	3.157	8.342	4.242	4.242	-1.731
		(0.004)	(0.045)		(0.093)	(0.082)	
Create time	20047	20047	20056	-3.952	20027	20012	3.227
		(0.185)	(2.309)		(3.291)	(3.294)	
Tenure	552	554	244	141.990	287	292	-1.232
		(0.171)	(1.661)		(2.389)	(2.580)	
Obs.	6,151,717	6,114,124	37,593		18,841	18,841	

Table 2.5. Descriptive statistics for Chapter II

	VARIABLES	N	mean	sd	min	max	1	2	3	4	5	6	7	8	9	10	11
1	User adoption	322,410	1.87	2.79	0	15.89											
2	Cross-platform complementor	322,410	0.57	0.50	0	1	0.02										
3	After	322,410	0.56	0.50	0	1	0.02	0.08									
4	Watchers	322,410	0.17	0.60	0	8.15	0.07	-0.01	-0.01								
5	Own dependencies	322,410	1.22	1.42	0	9.81	0.31	0.01	-0.02	0.00							
6	Relative platform size	322,410	4.91	10.91	0	44.73	0.06	-0.04	-0.07	0.00	0.15						
7	Pull requests	322,410	0.00	0.05	0	4.34	0.00	0.00	0.00	0.06	-0.01	-0.01					
8	Issues	322,410	0.06	0.34	0	6.54	0.04	0.00	-0.01	0.57	0.00	-0.01	0.09				
9	Forks	322,410	0.23	0.70	0	10.25	0.29	0.01	0.01	0.02	0.11	0.03	-0.01	0.01			
10	Commits	322,410	0.18	0.72	0	9.04	0.06	0.00	-0.02	0.42	0.03	-0.01	0.17	0.52	0.00		
11	Number of libraries	322,410	4.54	12.59	0	368	0.12	-0.01	-0.01	0.12	0.09	0.12	0.00	0.06	0.23	0.09	
12	Platform libraries	322,410	10.93	1.41	2.4	13.52	0.13	-0.06	-0.01	-0.01	0.33	0.32	-0.03	-0.04	0.07	-0.03	0.20

Table 2.6. OLS estimations of the effects of multihoming on user adoption

DV: log of dependent repositories	(1)	(2)	(3)	(4)	(5)	(6)
Cross-platform complementor (Treated)		0.059***	0.069***			
		(0.009)	(0.009)			
After (Post treatment)		0.159***	0.156***		0.077***	0.069***
		(0.009)	(0.009)		(0.006)	(0.005)
Cross-platform complementor X After (H1: β > 0)			0.119***			0.203***
			(0.018)			(0.006)
Watchers	0.192***	0.193***	0.193***	0.063***	0.063***	0.061***
	(0.010)	(0.010)	(0.010)	(0.007)	(0.007)	(0.007)
Own dependencies	0.533***	0.533***	0.532***	0.160***	0.159***	0.156***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Relative platform size	0.000	0.001+	0.001*	0.002	0.002	0.003
	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)
Pull requests	-0.117	-0.117	-0.119	-0.022	-0.022	-0.025
	(0.081)	(0.080)	(0.080)	(0.067)	(0.067)	(0.067)
Issues	0.028	0.028	0.028	0.014	0.014	0.013
	(0.019)	(0.019)	(0.019)	(0.011)	(0.011)	(0.011)
Forks	1.016***	1.014***	1.013***	0.136***	0.136***	0.134***
	(0.010)	(0.010)	(0.010)	(0.006)	(0.006)	(0.006)
Commits	0.123***	0.126***	0.126***	0.035***	0.035***	0.034***
	(0.008)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)
Number of libraries	0.006***	0.006***	0.006***	0.008***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Platform libraries	0.030***	0.030***	0.031***	0.800***	0.794***	0.791***
	(0.003)	(0.003)	(0.003)	(0.036)	(0.036)	(0.036)
Constant	1.281***	1.284***	1.276***	-9.982***	-9.501***	-9.486***
	(0.034)	(0.034)	(0.034)	(0.341)	(0.343)	(0.342)
Observations	322,410	322,410	322,410	322,410	322,410	322,410
R-squared	0.167	0.168	0.168	0.926	0.926	0.926
Lib fixed effect	NO	NO	NO	YES	YES	YES
Month fixed effect	NO	NO	NO	YES	YES	YES

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 2.7. OLS estimations of contingencies in multihoming and user adoption

DV: log of dependent repositories VARIABLES	(1) Contingency: Watchers (H2: $\beta > 0$)	(2) Contingency: Own dependencies (H3: $\beta < 0$)	(3) Contingency: Relative platform size (H4: $\beta < 0$)
After (Post treatment)	0.069***	0.071***	0.069***
	(0.005)	(0.005)	(0.005)
Cross-platform complementor X After (H: $\beta > 0$)	0.201***	0.215***	0.203***
	(0.006)	(0.006)	(0.006)
Cross-platform complementor X After X Contingency	0.046***	-0.030***	-0.002***
	(0.010)	(0.005)	(0.001)
Cross-platform complementor X Contingency	0.084***	-0.129***	-0.010*
	(0.014)	(0.013)	(0.004)
After X Contingency	-0.020***	0.039***	-0.001***
	(0.005)	(0.002)	(0.000)
Watchers	0.051***	0.059***	0.060***
	(0.007)	(0.007)	(0.007)
Own dependencies	0.155***	0.196***	0.155***
	(0.004)	(0.006)	(0.004)
Relative platform size	0.003	0.003	0.000
	(0.002)	(0.002)	(0.002)
Pull requests	-0.022	-0.015	-0.016
	(0.065)	(0.065)	(0.065)
Issues	0.011	0.012	0.013
	(0.010)	(0.010)	(0.010)
Forks	0.129***	0.127***	0.129***
	(0.006)	(0.006)	(0.006)
Commits	0.035***	0.035***	0.035***
	(0.005)	(0.005)	(0.005)
Number of libraries	0.008***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)
Platform libraries	0.801***	0.813***	0.810***
	(0.036)	(0.036)	(0.036)
Constant	-9.564***	-9.670***	-9.640***
	(0.341)	(0.341)	(0.342)
Observations	322,410	322,410	322,410
R-squared	0.925	0.925	0.925
Lib fixed effect	YES	YES	YES
Month fixed effect	YES	YES	YES

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 2.8. Model sensitivity to time window specifications

	Time window: 3-month before and after				Time window: 12-month before and after			
DV: log of dependent repositories	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After (Post treatment)	0.080***	0.080***	0.083***	0.078***	0.048***	0.050***	0.047***	0.050***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Cross-platform complementor X After (H1: $\beta > 0$)	0.175***	0.174***	0.176***	0.175***	0.247***	0.246***	0.285***	0.248***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Cross-platform complementor X After X Contingency		0.037**	-0.023***	-0.003***		0.042***	-0.046***	-0.002***
		(0.012)	(0.005)	(0.001)		(0.009)	(0.005)	(0.001)
Cross-platform complementor X Contingency		0.087***	-0.051**	-0.008		0.075***	-0.211***	-0.013***
		(0.018)	(0.019)	(0.008)		(0.011)	(0.009)	(0.002)
After X Contingency		-0.006	0.022***	-0.001***		-0.037***	0.063***	-0.001***
		(0.006)	(0.003)	(0.000)		(0.005)	(0.002)	(0.000)
Watchers	0.044***	0.034***	0.044***	0.044***	0.084***	0.078***	0.082***	0.084***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.006)	(0.006)	(0.006)	(0.006)
Own dependencies	0.121***	0.120***	0.140***	0.121***	0.203***	0.201***	0.262***	0.202***
	(0.007)	(0.007)	(0.009)	(0.007)	(0.003)	(0.003)	(0.005)	(0.003)
Relative platform size	0.011**	0.010**	0.010*	0.007+	0.002	0.001	0.001	-0.000
	(0.004)	(0.004)	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
Pull requests	-0.068	-0.071	-0.063	-0.066	-0.053	-0.053	-0.049	-0.044
	(0.080)	(0.080)	(0.080)	(0.080)	(0.050)	(0.049)	(0.049)	(0.049)
Issues	0.024	0.019	0.022	0.022	0.027**	0.025**	0.027**	0.028**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.009)	(0.009)	(0.009)	(0.009)
Forks	0.109***	0.106***	0.106***	0.106***	0.211***	0.202***	0.199***	0.202***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.005)	(0.005)	(0.005)	(0.005)
Commits	0.029***	0.028***	0.029***	0.029***	0.044***	0.043***	0.044***	0.044***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Number of libraries	0.010***	0.009***	0.009***	0.009***	0.008***	0.007***	0.007***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Platform libraries	0.939***	0.947***	0.964***	0.966***	0.668***	0.673***	0.684***	0.676***
	(0.082)	(0.082)	(0.082)	(0.082)	(0.018)	(0.018)	(0.018)	(0.018)
Constant	-10.915***	-10.981***	-11.146***	-11.152***	-8.035***	-8.053***	-8.159***	-8.074***
	(0.753)	(0.750)	(0.752)	(0.752)	(0.188)	(0.186)	(0.186)	(0.187)
Observations	158,599	158,599	158,599	158,599	616,724	616,724	616,724	616,724
R-squared	0.953	0.952	0.952	0.952	0.891	0.888	0.889	0.888
Lib fixed effect	YES	YES	YES	YES	YES	YES	YES	YES
Month fixed effect	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses, *** $p<0.001$, ** $p<0.01$, * $p<0.05$, + $p<0.1$

Table 2.9. Tests of model sensitively to low usage complementary technologies

DV: log of dependent repositories	(1)	(2) Contingency: Watchers (H2: β > 0)	(3) Contingency: Own dependencies (H3: β < 0)	(4) Contingency: Relative platform size (H4: β < 0)
After (Post treatment)	0.111***	0.110***	0.097***	0.110***
	(0.007)	(0.007)	(0.007)	(0.007)
Cross-platform complementor X After (H1: β > 0)	0.196***	0.198***	0.235***	0.202***
	(0.008)	(0.008)	(0.008)	(0.008)
Cross-platform complementor X After X Contingency		0.027*	-0.121***	-0.000***
		(0.014)	(0.008)	(0.000)
Cross-platform complementor X Contingency		0.056**	-0.094***	-0.000
		(0.019)	(0.014)	(0.000)
After X Contingency		-0.003	0.054***	0.000
		(0.007)	(0.004)	(0.000)
Watchers	0.071***	0.066***	0.068***	0.070***
	(0.010)	(0.010)	(0.010)	(0.010)
Own dependencies	0.212***	0.212***	0.242***	0.212***
	(0.007)	(0.007)	(0.007)	(0.007)
Relative platform size	-0.044	-0.033	-0.033	-0.033
	(0.083)	(0.079)	(0.079)	(0.079)
Pull requests	0.008	0.005	0.008	0.008
	(0.014)	(0.014)	(0.014)	(0.014)
Issues	0.138***	0.133***	0.131***	0.133***
	(0.009)	(0.009)	(0.009)	(0.009)
Forks	0.029***	0.030***	0.030***	0.030***
	(0.006)	(0.006)	(0.006)	(0.006)
Commits	0.017***	0.017***	0.016***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
Number of libraries	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Platform libraries	-0.158***	-0.158***	-0.144***	-0.159***
	(0.009)	(0.009)	(0.010)	(0.009)
Constant	-0.669***	-0.647***	-0.939***	-0.635**
	(0.195)	(0.194)	(0.195)	(0.195)
Observations	195,937	195,937	195,937	195,937
R-squared	0.935	0.933	0.934	0.933
Lib fixed effect	YES	YES	YES	YES
Month fixed effect	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1,
Sample excluding complementors without pre-treatment user adoption

Table 3.1. Examples of open source-based ventures with VC investment

Venture name	Technology type	Total VC investment	Round	Latest round time	Collaborations*	Number of repositories
Docker	Cloud computing	91.9 M	E	Oct-17	19712	146**
MongoDB	Database	306.1 M	F***	Jan-15	1365	125
Elastic	Cloud computing	162 M	D***	Jul- 16	20230	403
Mapbox	Mapping service	227.2 M	C	Oct-17	6950	801
Confluent	Streaming service	205.9 M	D	Jan-19	1616	88
NPM	Package manager	10.6 M	A	Apr-15	2487	267

* Only for the repository of the venture with highest number of collaboration requests, up to April 2019

** In 2018, Docker spanned out its most popular repository as Moby as a separate account, the statistics is based on Docker & Moby combined

*** Last round before IPO

Table 3.2. Summary of measures in Chapter III

CONSTRUCTS	*VARIABLE*	*MEASURES*
Dependent variable		
Value creation	Venture capital investment*	Total $ amount of venture capital invested to a venture with GitHub repositories
Independent and contingency variables		
Crowd collaboration	Pull requests*	Number of pull requested (request of change in source codes) opened by the crowd contributors to a venture's repositories (log transformed)
Knowledge disclosure	Pushes*	Number of pushes (updates on source code) by a venture to its repositories
Knowledge breadth	Programming language diversity	Herfindahl index of the concentration of programming languages used in a ventures' repositories
Control variables		
User interests/base	Watchers*	Number of people who subscribed the updates of a ventures' repositories
Innovation quality	Issues*	Number of bugs and errors in the ‘issue’ section of a ventures' GitHub repositories
Diffusion	Forks*	Number of developers copied the source code files of a venture's repositories to the directory under their own GitHub individual account
Venture knowledge	Programming languages	Number of programming languages used by a venture's GitHub repositories
Ventures’ technological capabilities	Number of repositories	Number of GitHub repositories initiated by a venture
Tenure		Number of years since a venture opened its first repository on GitHub

* Log transformed

Table 3.3. Probit estimation on the likelihood of crowd collaborations

	(1)	
DV	Crowd collaboration (treated)	
Pushes	0.163	•••
	(0.001)	
Knowledge breadth	0.226	•••
	(0.005)	
Forks	0.525	•••
	(0.002)	
Issues	0.183	•••
	(0.001)	
Watchers	0.058	•••
	(0.002)	
Programming languages	0.019	•••
	(0.002)	
Number of repositories	0.052	•••
	(0.003)	
Tenure	-0.159	•••
	(0.001)	
Constant	-1.321	•••
	(0.026)	
Observations	2,942,510	
Quarter fixed effects	Yes	

••• p<0.001, •• p<0.01, • p<0.05, † p<0.1; Robust standard errors in parentheses, two-tailed tests

Table 3.4. *t*-Tests of open source-based ventures before and after matching

	Unmatched sample				Matched sample		
VARIABLES	All	Control	Treated	t-value	Control	Treated	t-value
VC investment (original amount $ thousand)	43.87	35.70	171.27	-19.23	40.58	232.22	-5.49
	(1.68)	(1.35)	(18.18)		(10.52)	(32.75)	
Pushes	2.43	2.35	3.76	-3.50E+02	3.46	3.39	8.41
	(0.00)	(0.00)	(0.00)		(0.01)	(0.01)	
Knowledge breadth	0.36	0.36	0.51	-1.70E+02	0.52	0.50	10.47
	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Forks	0.32	0.27	1.14	-5.60E+02	0.60	0.65	-11.66
	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Issues	0.27	0.22	1.02	-4.20E+02	0.51	0.51	-0.46
	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Watchers	0.43	0.38	1.18	-4.10E+02	0.68	0.72	-7.54
	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Programming languages	1.60	1.58	2.02	-1.50E+02	1.89	1.97	-9.47
	(0.00)	(0.00)	(0.00)		(0.01)	(0.01)	
Number of repositories	1.32	1.30	1.73	-2.50E+02	1.66	1.66	-0.71
	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Tenure	6.42	6.68	2.21	338.89	3.36	3.51	-6.29
	(0.00)	(0.00)	(0.01)		(0.02)	(0.02)	
Obs.	2,952,748	2,781,745	171,003		64,590	64,590	

Table 3.5. Descriptive statistics for Chapter III

	Variable	Mean	SD	Min.	Max.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1)	Venture capital investment	0.09	0.90	0	15.60									
(2)	Crowd collaboration	0.18	0.60	0	7.67	0.07								
(3)	Pushes	1.03	1.55	0	12.25	0.04	0.49							
(4)	Knowledge breadth	0.47	0.34	0	1.00	0.04	0.09	0.16						
(5)	Forks	0.92	1.18	0	11.17	0.06	0.27	0.18	0.12					
(6)	Issues	0.79	0.98	0	9.53	0.06	0.31	0.15	0.12	0.68				
(7)	Watchers	0.66	1.18	0	9.98	0.01	0.29	0.22	0.08	0.43	0.36			
(8)	Programming languages	1.69	0.80	0.69	4.75	0.09	0.20	0.27	0.55	0.32	0.34	0.18		
(9)	Number of repositories	6.73	5.28	0	26	0.06	0.00	-0.20	0.07	0.30	0.32	0.18	0.24	
(10)	Tenure	2.13	1.61	1	30	0.11	0.17	0.20	0.47	0.28	0.30	0.15	0.70	0.25

Table 3.6. OLS estimations on the effect crowd collaboration on venture capital investment to open source-based ventures

DV: log of venture capital investment	(1)		(2)		(3)		(4)		(5)	
Crowd collaboration			0.068	•••	0.077	•••	0.065	•••	0.076	•••
			(0.008)		(0.008)		(0.008)		(0.008)	
Crowd collaboration X Pushes					-0.012	•••			-0.013	•••
					(0.002)				(0.002)	
Crowd collaboration X Knowledge breadth							0.042	•••	0.055	•••
							(0.009)		(0.009)	
Pushes	0.003	•••	0.000		-0.008	•••	0.000		-0.008	•••
	(0.000)		(0.000)		(0.001)		(0.000)		(0.001)	
Knowledge breadth	-0.053	•••	-0.051	•••	-0.052	•••	-0.066	•••	-0.063	•••
	(0.006)		(0.006)		(0.006)		(0.007)		(0.007)	
Watchers	0.024	•••	0.024	•••	0.025	•••	0.024	•••	0.025	•••
	(0.002)		(0.002)		(0.002)		(0.002)		(0.002)	
Forks	0.014	•••	0.011	•••	0.011	•••	0.011	•••	0.011	•••
	(0.002)		(0.002)		(0.002)		(0.002)		(0.002)	
Issues	-0.025	•••	-0.027	•••	-0.026	•••	-0.027	•••	-0.026	•••
	(0.001)		(0.001)		(0.001)		(0.001)		(0.001)	
Number of repositories	0.047	•••	0.048	•••	0.050	•••	0.049	•••	0.051	•••
	(0.005)		(0.005)		(0.005)		(0.005)		(0.005)	
Tenure	-0.001		0.000		-0.001		0.000		-0.001	
	(0.001)		(0.001)		(0.002)		(0.001)		(0.001)	
Programming languages	0.046	•••	0.045	•••	0.045	•••	0.045	•••	0.045	•••
	(0.002)		(0.002)		(0.002)		(0.002)		(0.002)	
Constant	-0.076	•••	-0.090	•••	-0.077	•••	-0.094	•••	-0.081	•••
	(0.019)		(0.019)		(0.019)		(0.018)		(0.019)	
Observations	1,006,412		1,006,412		1,006,412		1,006,412		1,006,412	
R-squared	0.842		0.842		0.842		0.842		0.842	
Owner fixed effects	YES		YES		YES		YES		YES	
Quarter fixed effects	YES		YES		YES		YES		YES	
Treatment X Quarter fixed effects	NO		YES		YES		YES		YES	
Contingency X Quarter fixed effects	NO		NO		YES		YES		YES	

••• $p<0.001$, •• $p<0.01$, • $p<0.05$, † $p<0.1$; Robust standard errors in parentheses, two-tailed tests

Table 3.7. Tests of model sensitivity to fixed effects specifications, observation periods and log transformation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					Observation period (unit: quarter)			Non-logged knowledge disclosure	Non-logged knowledge disclosure
	Final model	No fixed effects	Owner quarter fixed effects	Fixed effects with XTREG	(-6, 6)	(-10,10)	(-2,8)		
Crowd collaboration	0.076 •••	0.287 •••	0.076 •••	0.076 •••	0.073 •••	0.073 •••	0.051 •••	0.073 •••	0.070 •••
	(0.008)	(0.009)	(0.008)	(0.012)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Crowd collaboration X Pushes	-0.013 •••	-0.037 •••	-0.013 •••	-0.013 •••	-0.011 •••	-0.015 •••	-0.006 •••	-0.000 •••	-0.000 •••
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)
Crowd collaboration X Knowledge breadth	0.055 •••	0.149 •••	0.055 •••	0.055 •••	0.040 •••	0.059 •••	0.039 •••		0.049 •••
	(0.009)	(0.015)	(0.009)	(0.014)	(0.009)	(0.009)	(0.009)		(0.009)
Pushes	-0.008 •••	0.007 •••	-0.008 •••	-0.008 •••	-0.008 •••	-0.009 •••	-0.006 •••	-0.000 •••	-0.000 •••
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Knowledge breadth	-0.063 •••	-0.068 •••	-0.063 •••	-0.063 •••	-0.054 •••	-0.072 •••	-0.057 •••	-0.052 •••	-0.066 •••
	(0.007)	(0.003)	(0.007)	(0.012)	(0.007)	(0.006)	(0.008)	(0.006)	(0.007)
Watchers	0.025 •••	0.012 •••	0.025 •••	0.025 •••	0.021 •••	0.026 •••	0.020 •••	0.024 •••	0.024 •••
	(0.002)	(0.001)	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Forks	0.011 •••	-0.000	0.011 •••	0.011 ••	0.011 •••	0.012 •••	0.013 •••	0.011 •••	0.011 •••
	(0.002)	(0.001)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Issues	-0.026 •••	-0.025 •••	-0.026 •••	-0.026 •••	-0.023 •••	-0.029 •••	-0.019 •••	-0.026 •••	-0.027 •••
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Number of repositories	0.051 •••	0.030 •••	0.051 •••	0.051 •••	0.049 •••	0.051 •••	0.042 •••	0.048 •••	0.049 •••
	(0.005)	(0.002)	(0.005)	(0.009)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
Tenure	-0.001	0.006 •••	-0.001	-0.001	0.000	-0.001	-0.001	0.001	0.001
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Programming languages	0.045 •••	0.042 •••	0.045 •••	0.045 •••	0.039 •••	0.049 •••	0.039 •••	0.045 •••	0.045 •••
	(0.002)	(0.001)	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Constant	-0.081 •••	-0.100 •••	-0.081 •••	-0.081 •••	-0.077 •••	-0.093 •••	-0.052 •••	-0.102 •••	-0.105 •••
	(0.019)	(0.003)	(0.019)	(0.015)	(0.019)	(0.019)	(0.015)	(0.019)	(0.019)
Observations	1,006,412	1,006,412	1,006,412	1,006,412	853,268	1,117,578	781,962	1,006,412	1,006,412
R-squared	0.842	0.015	0.842	0.021	0.859	0.830	0.897	0.842	0.842
Owner fixed effects	YES	NO	YES	YES	YES	YES	YES	YES	YES
Quarter fixed effects	YES	NO	YES	YES	YES	YES	YES	YES	YES
Treatment X Quarter fixed effects	YES	NO	NO	YES	YES	YES	YES	YES	YES
Contengy X Quarter fixed effects	YES	NO	NO	YES	YES	YES	YES	YES	YES

••• $p<0.001$, •• $p<0.01$, • $p<0.05$, † $p<0.1$; Robust standard errors in parentheses, two-tailed tests

Only knowledge disclosure is tested in non-log transformed value due to its moderate correlation with crowd collaboation in Model 8 and Model 9

*DV: venture capital investment

Table 3.8. Testing mechanisms: OLS estimations on the effect of collation on lead users' learning interest (Forks) and user awareness (Watchers)

	(1)	(2)	(3)	(4)
DV	Forks		Watchers	
Crowd collaboration	0.195 •••	0.263 •••	0.190 •••	0.199 •••
	(0.003)	(0.004)	(0.007)	(0.007)
Crowd collaboration X Pushes		-0.032 •••		-0.013 •••
		(0.002)		(0.002)
Crowd collaboration X Knowledge breadth		0.077 •••		0.053 •••
		(0.008)		(0.009)
Pushes	0.006 •••	0.029 •••	0.019 •••	0.041 •••
	(0.000)	(0.001)	(0.000)	(0.002)
Knowledge breadth	-0.068 •••	-0.029 •••	-0.050 •••	-0.049 •••
	(0.004)	(0.005)	(0.005)	(0.006)
Watchers	0.406 •••	0.403 •••		
	(0.002)	(0.002)		
Forks			0.434 •••	0.433 •••
			(0.002)	(0.002)
Issues	0.095 •••	0.094 •••	0.170 •••	0.169 •••
	(0.001)	(0.001)	(0.002)	(0.002)
Number of repositories	0.211 •••	0.201 •••	0.210 •••	0.204 •••
	(0.003)	(0.003)	(0.004)	(0.004)
Tenure	0.021 •••	0.027 •••	0.015 •••	0.019 •••
	(0.001)	(0.001)	(0.001)	(0.002)
Programming languages	0.025 •••	0.024 •••	0.034 •••	0.034 •••
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.137 •••	-0.171 •••	0.339 •••	0.303 •••
	(0.016)	(0.018)	(0.018)	(0.020)
Observations	1,006,412	1,006,412	1,006,412	1,006,412
R-squared	0.928	0.928	0.940	0.940
Owner fixed effects	YES	YES	YES	YES
Quarter fixed effects	YES	YES	YES	YES
Treatment X Quarter fixed effects	YES	YES	YES	YES
Contingency X Quarter fixed effects	NO	YES	NO	YES

••• p<0.001, •• p<0.01, • p<0.05, † p<0.1; Robust standard errors in parentheses, two-tailed tests

Table 3.9. Alternative explanations of crowd collaboration as knowledge sourcing: the effects of collaboration accepted by ventures

DV: venture capital investment	(1)		(2)		(3)	
Collaboration accepted (Merged pull requests)	0.017	••	-0.003		-0.000	
	(0.006)		(0.007)		(0.007)	
Crowd collaboration			0.069	•••	0.076	•••
			(0.009)		(0.009)	
Crowd collaboration X Pushes					-0.015	•••
					(0.002)	
Crowd collaboration X Knowledge breadth					0.053	•••
					(0.009)	
Pushes	0.002	•••	0.000		-0.008	•••
	(0.000)		(0.000)		(0.001)	
Knowledge breadth	-0.053	•••	-0.051	•••	-0.063	•••
	(0.006)		(0.006)		(0.007)	
Watchers	0.024	•••	0.024	•••	0.025	•••
	(0.002)		(0.002)		(0.002)	
Forks	0.012	•••	0.011	•••	0.011	•••
	(0.002)		(0.002)		(0.002)	
Issues	-0.026	•••	-0.027	•••	-0.026	•••
	(0.001)		(0.001)		(0.001)	
Number of repositories	0.046	•••	0.047	•••	0.050	•••
	(0.005)		(0.005)		(0.005)	
Tenure	-0.001		0.000		-0.001	
	(0.001)		(0.001)		(0.001)	
Programming languages	0.045	•••	0.045	•••	0.045	•••
	(0.002)		(0.002)		(0.002)	
Constant	-0.081	•••	-0.092	•••	-0.083	•••
	(0.018)		(0.019)		(0.019)	
Observations	1,006,412		1,006,412		1,006,412	
R-squared	0.842		0.842		0.842	
Owner fixed effects	YES		YES		YES	
Quarter fixed effects	YES		YES		YES	
Treatment X Quarter fixed effects	YES		YES		YES	
Contingency X Quarter fixed effects	NO		NO		YES	

••• $p<0.001$, •• $p<0.01$, • $p<0.05$, † $p<0.1$; Robust standard errors in parentheses, two-tailed tests

LIST OF FIGURES

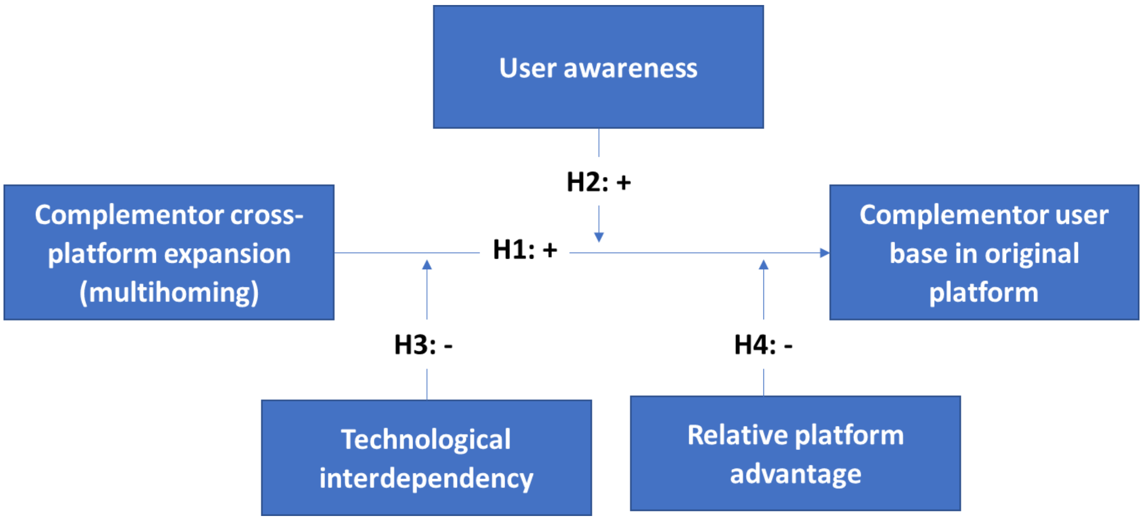
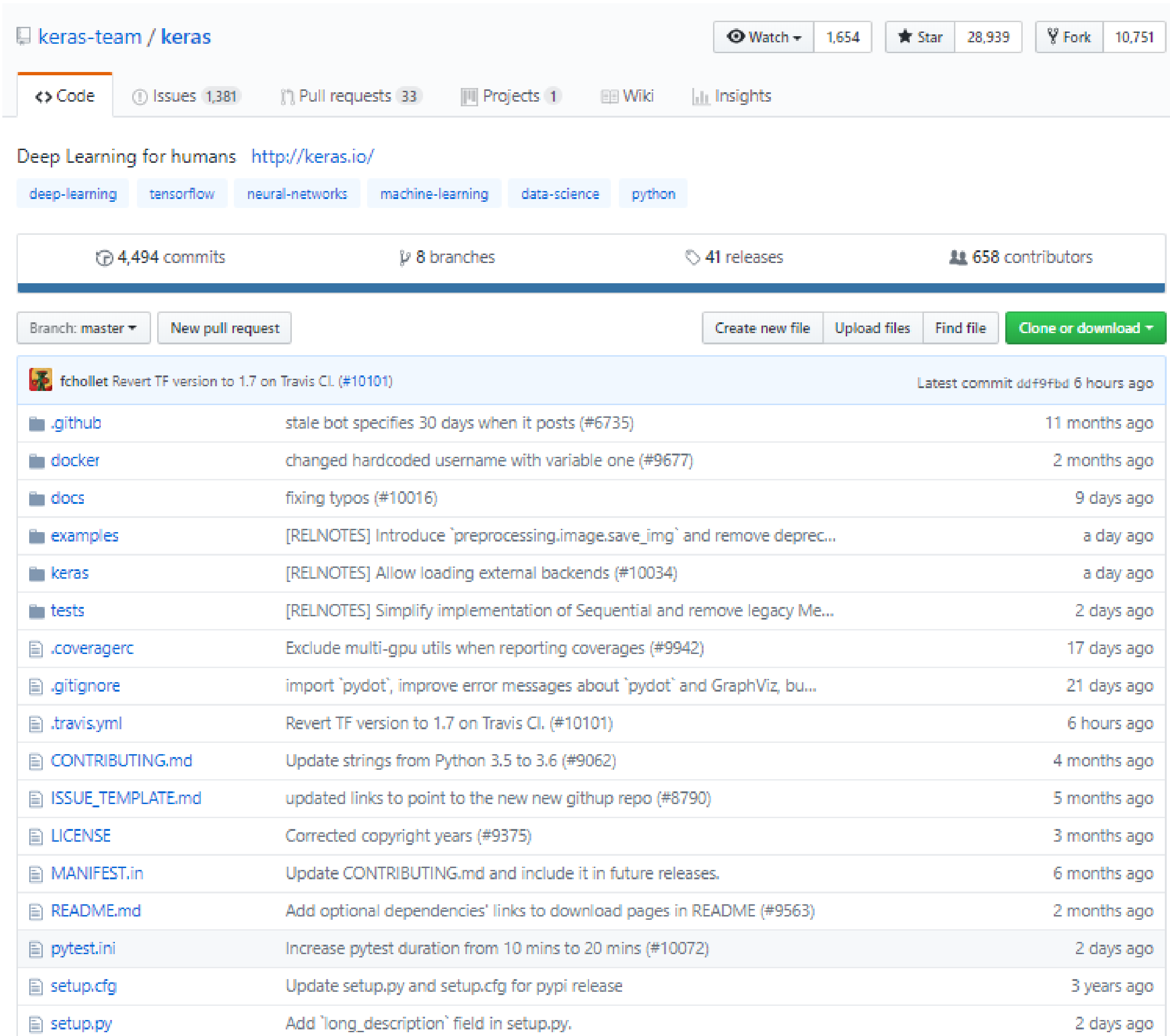
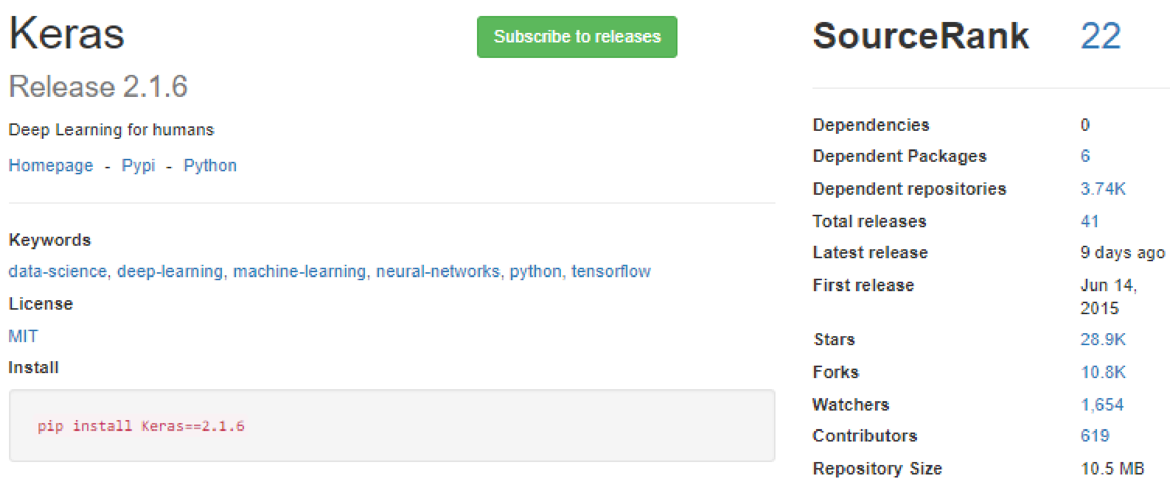


Figure 2.2A. GitHub repositories of a package library – example of Keras



Keras is a deep learning library initially developed for Python through the package manager Pypi. The illustration shows the repository information of Keras distributed through PyPi on Github.
Source: https://github.com/keras-team/keras

Figure 2.2B. Program package information on libraries.io – example of Keras



Keras is a deep learning library initially developed for Python through the package manager Pypi. The illustration shows the repository information of Keras distributed through PyPi on libareris.io. The details of source rank in: https://docs.libraries.io/overview.html#sourcerank

Source: https://libraries.io/pypi/Keras

Figure 2.3. Development and distribution process of open source libraries

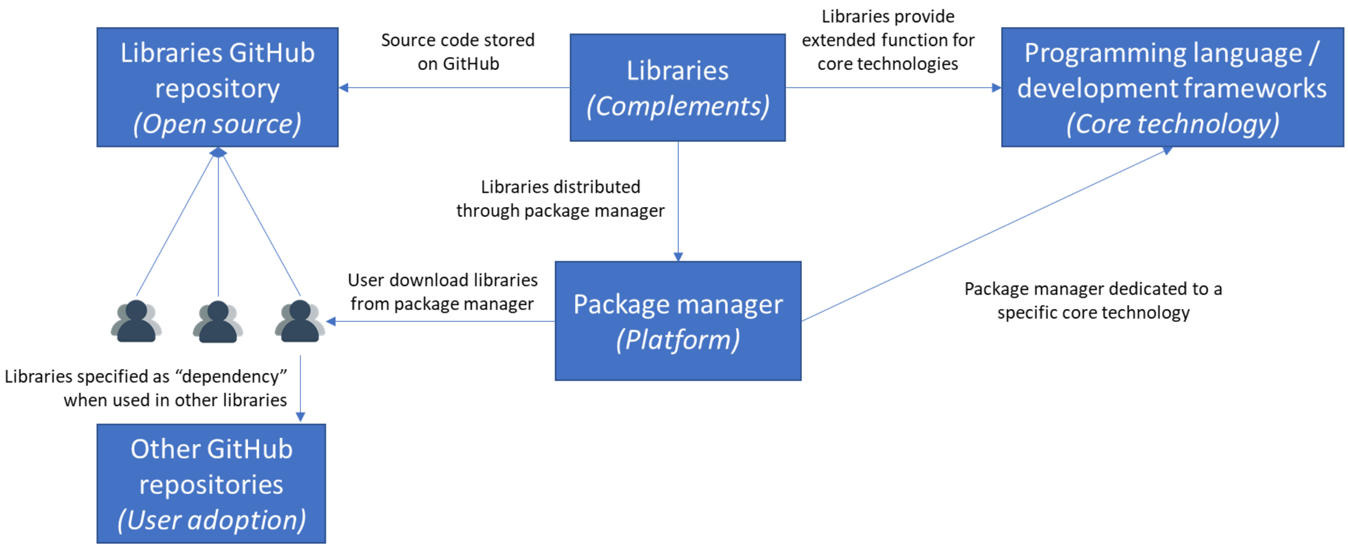


Figure 2.4. Growth of libraries and usage by month

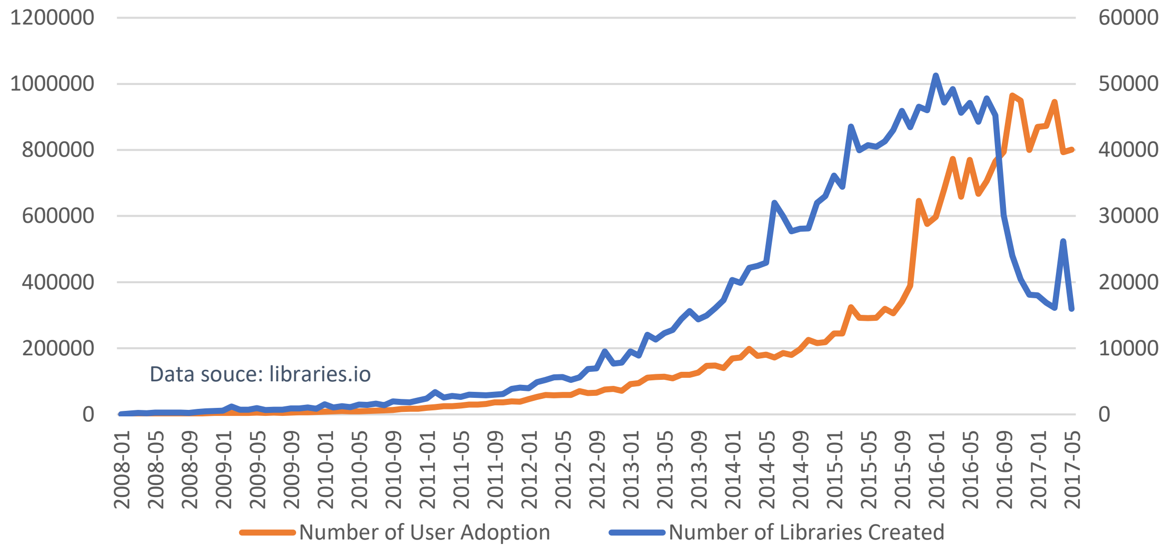


Figure 2.5. Sample package manager platforms and platform size by number of libraries (complements)

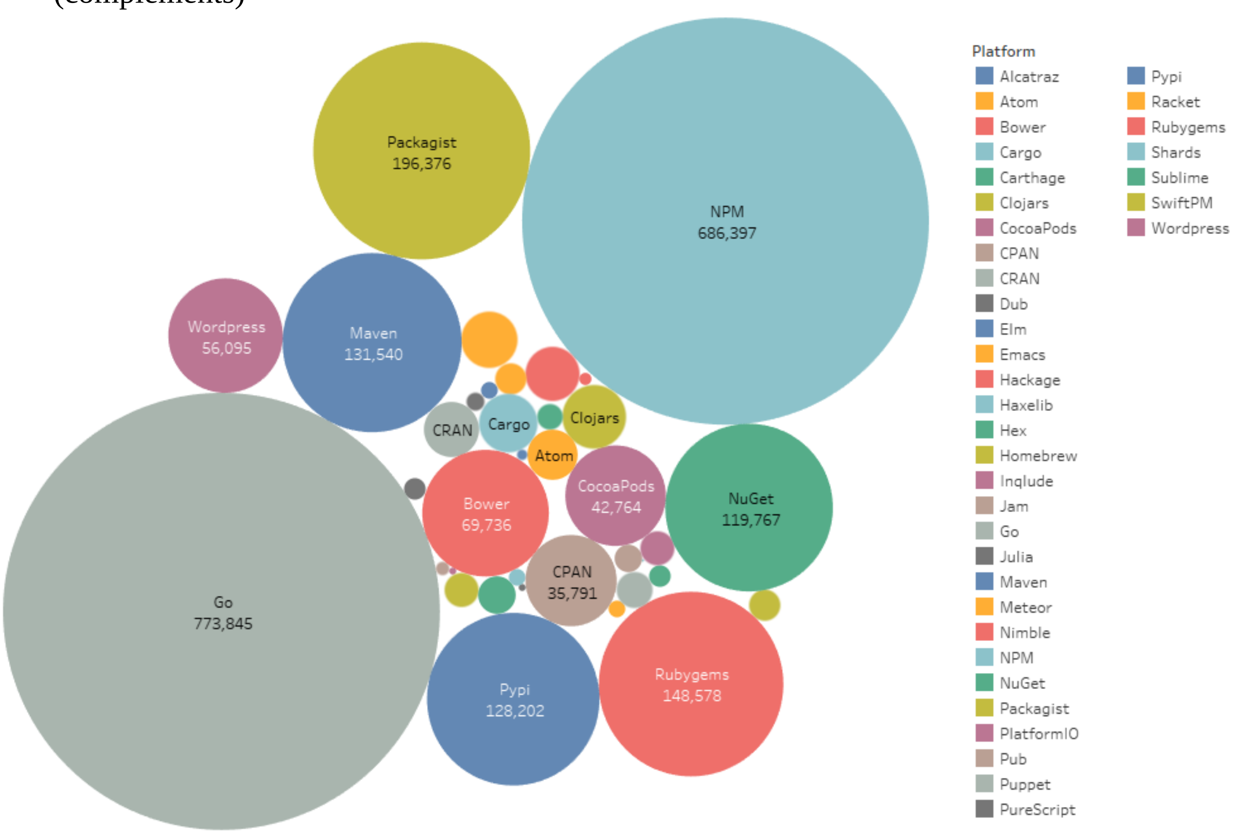


Figure 2.6. Percentage of multihoming complementors by platform

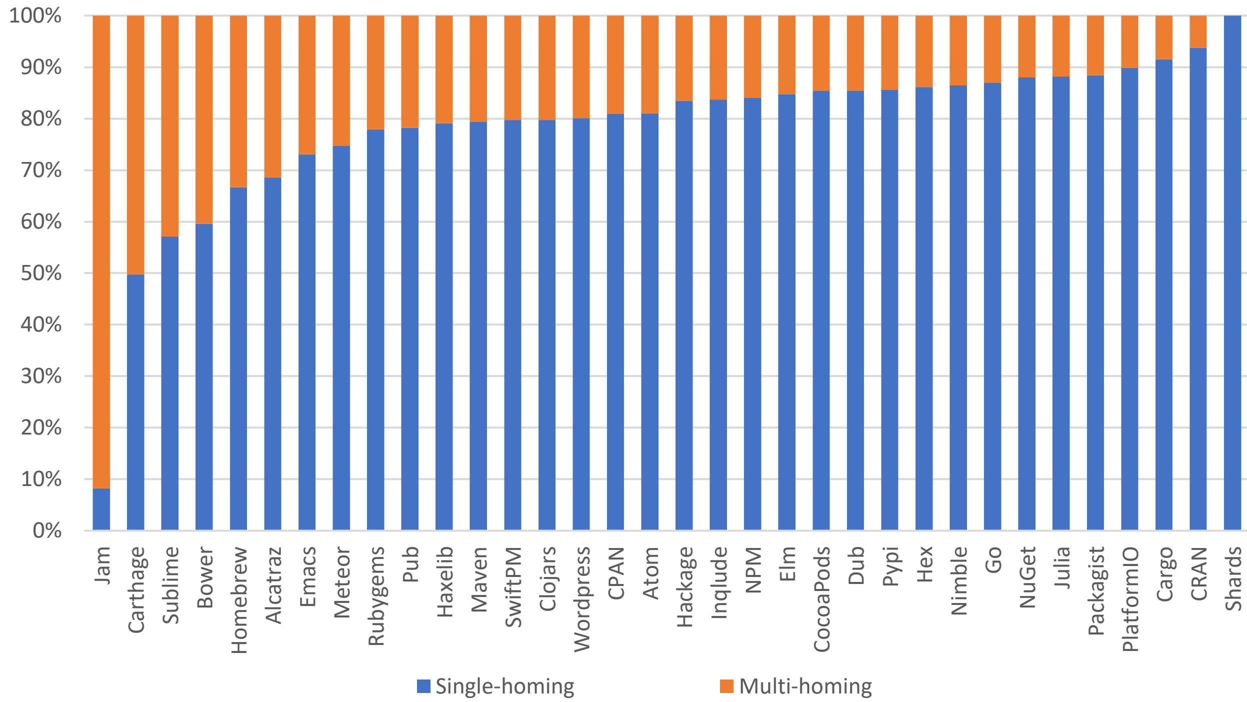


Figure 2.7. Distribution of the likelihood of complementor multihoming before and after matching

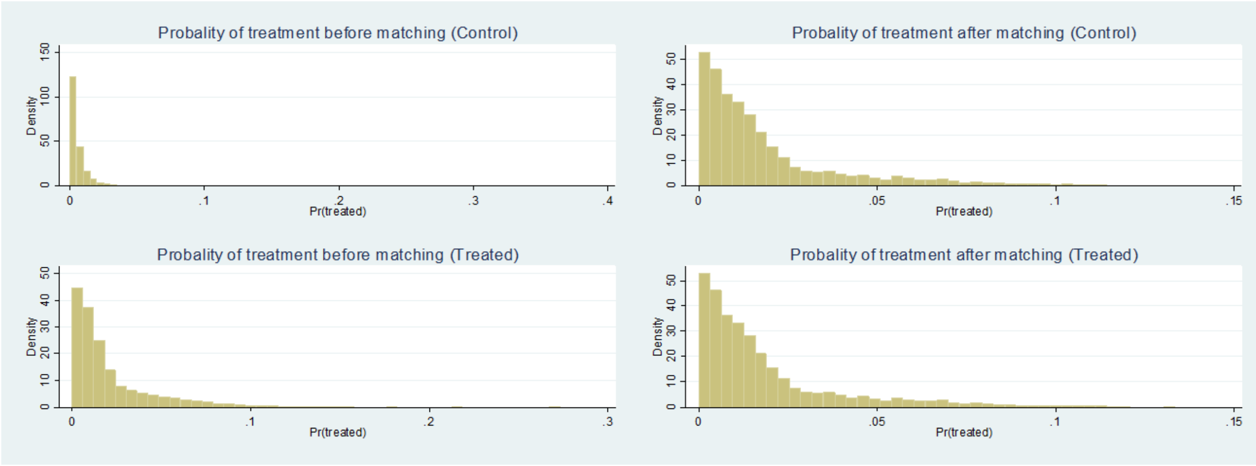


Figure 2.8. Comparison of user adoption pre-post treatment between treated and control complementors

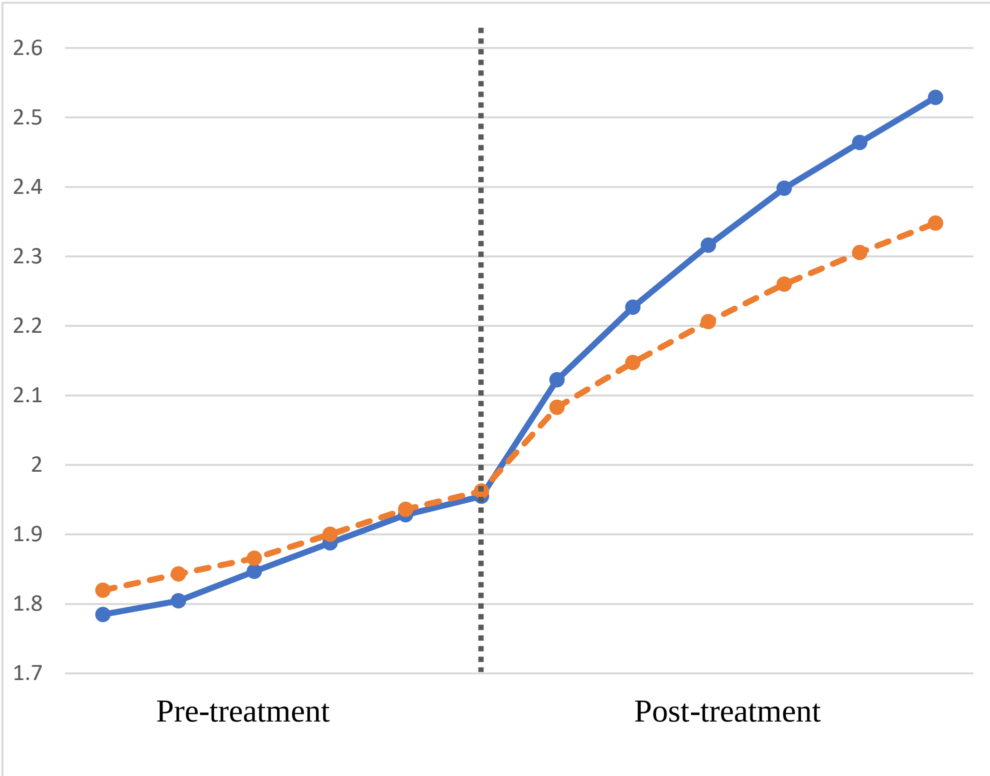


Figure 3.1. Summary of hypotheses (Chapter III)

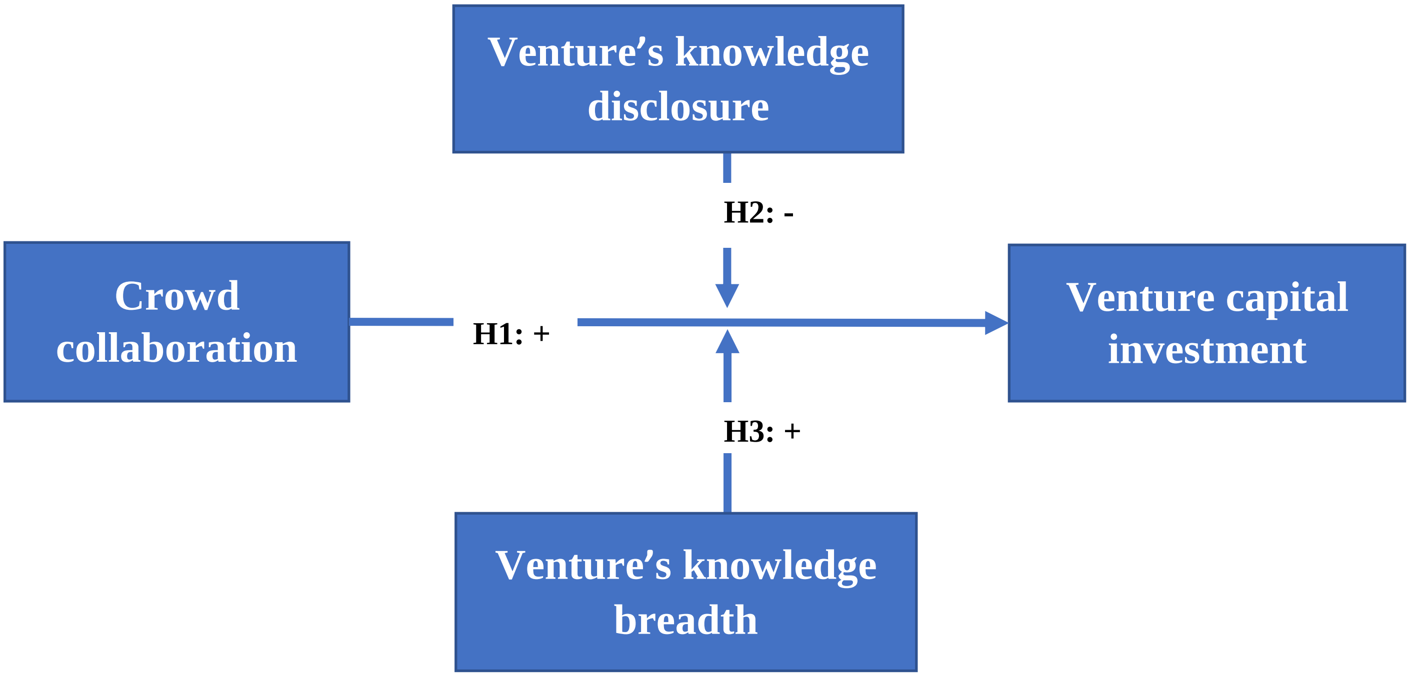


Figure 3.2. Venture capital investment to open source-based ventures by quarter

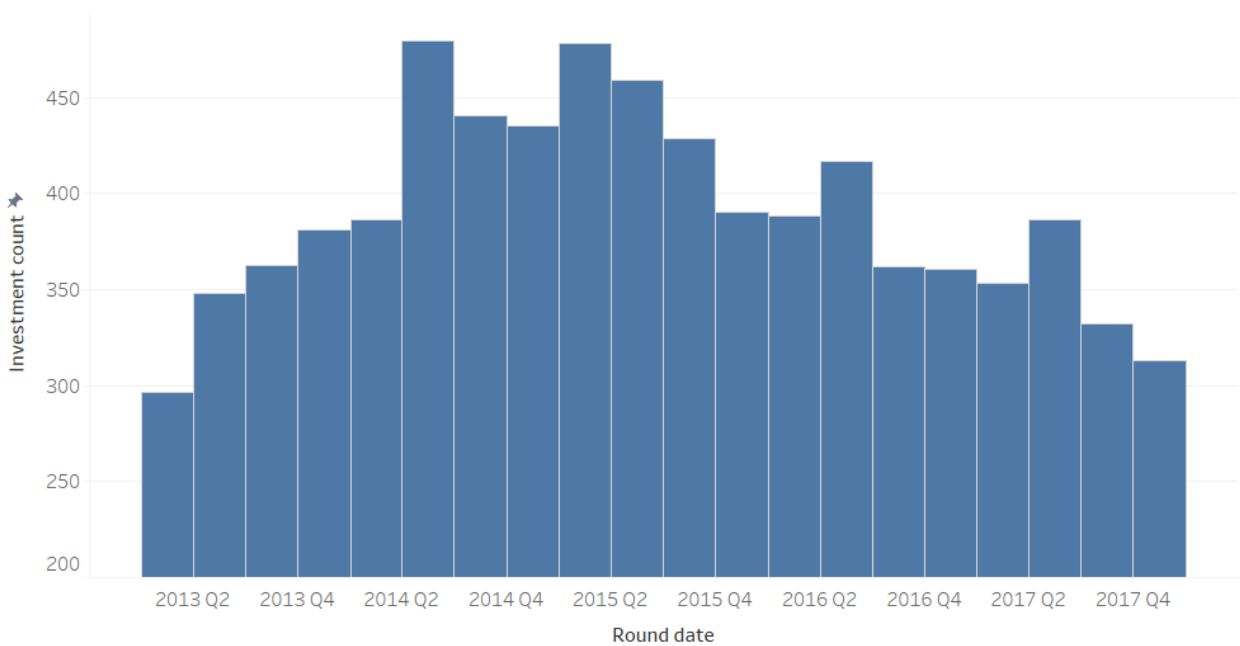


Figure 3.3. The crowd collaboration process on GitHub

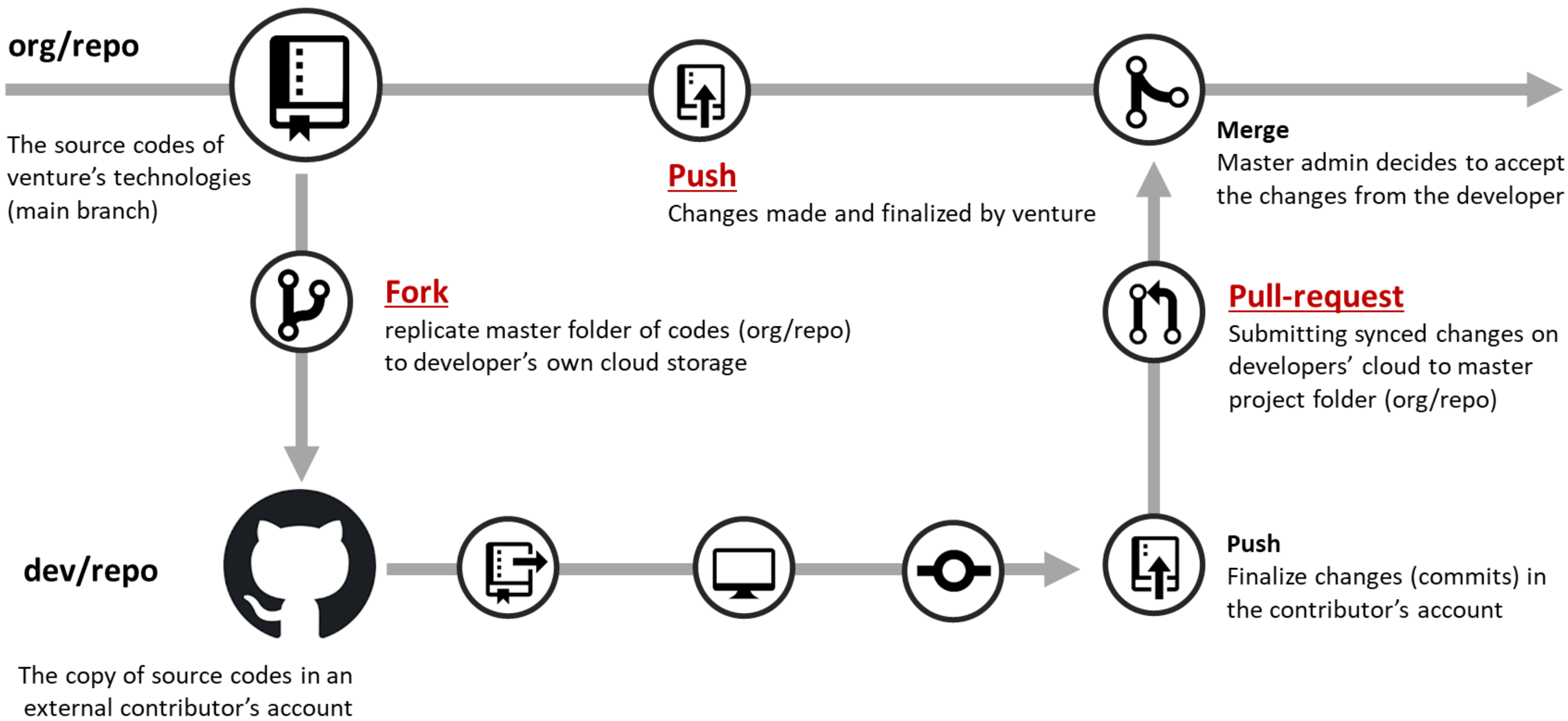
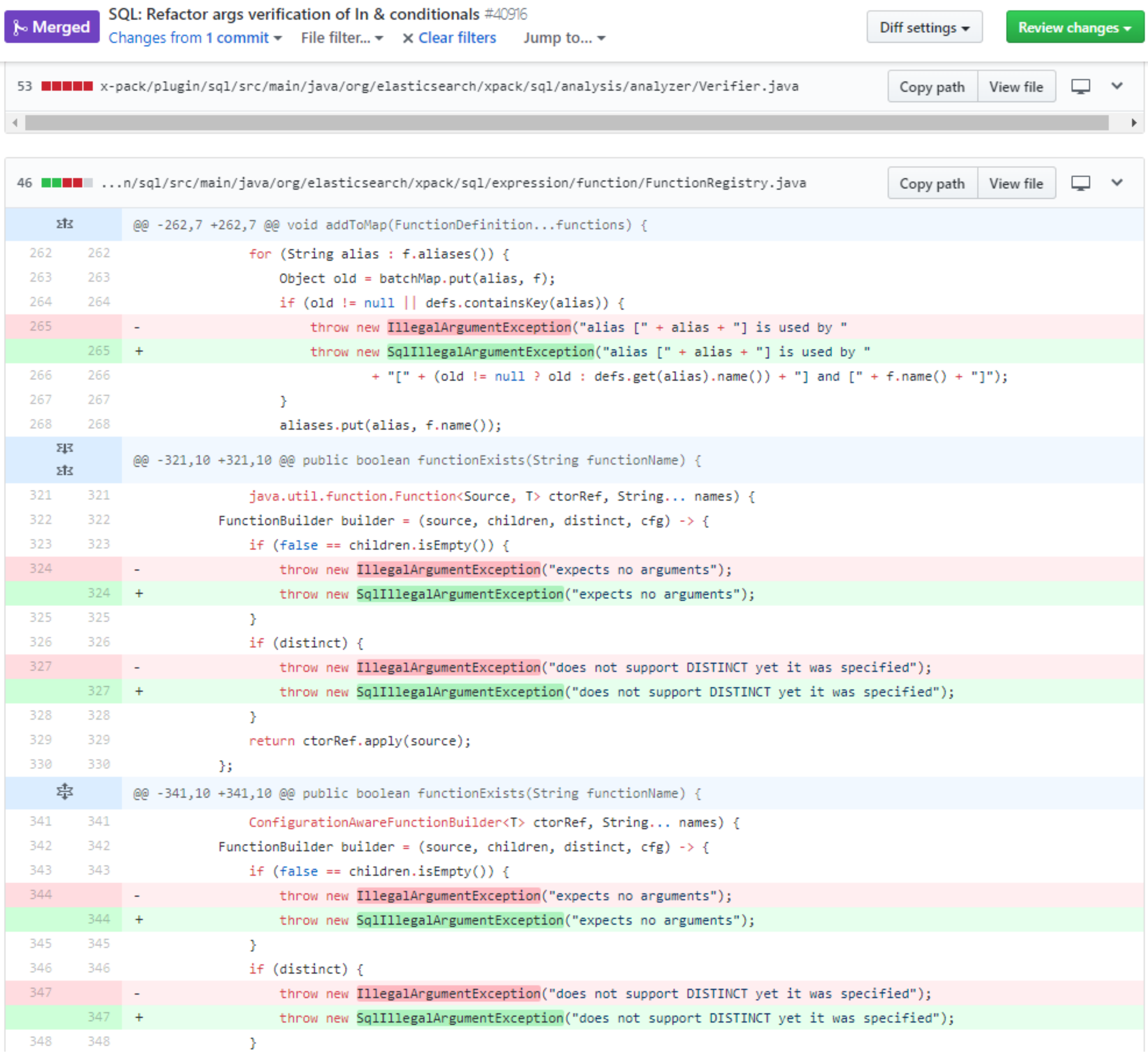


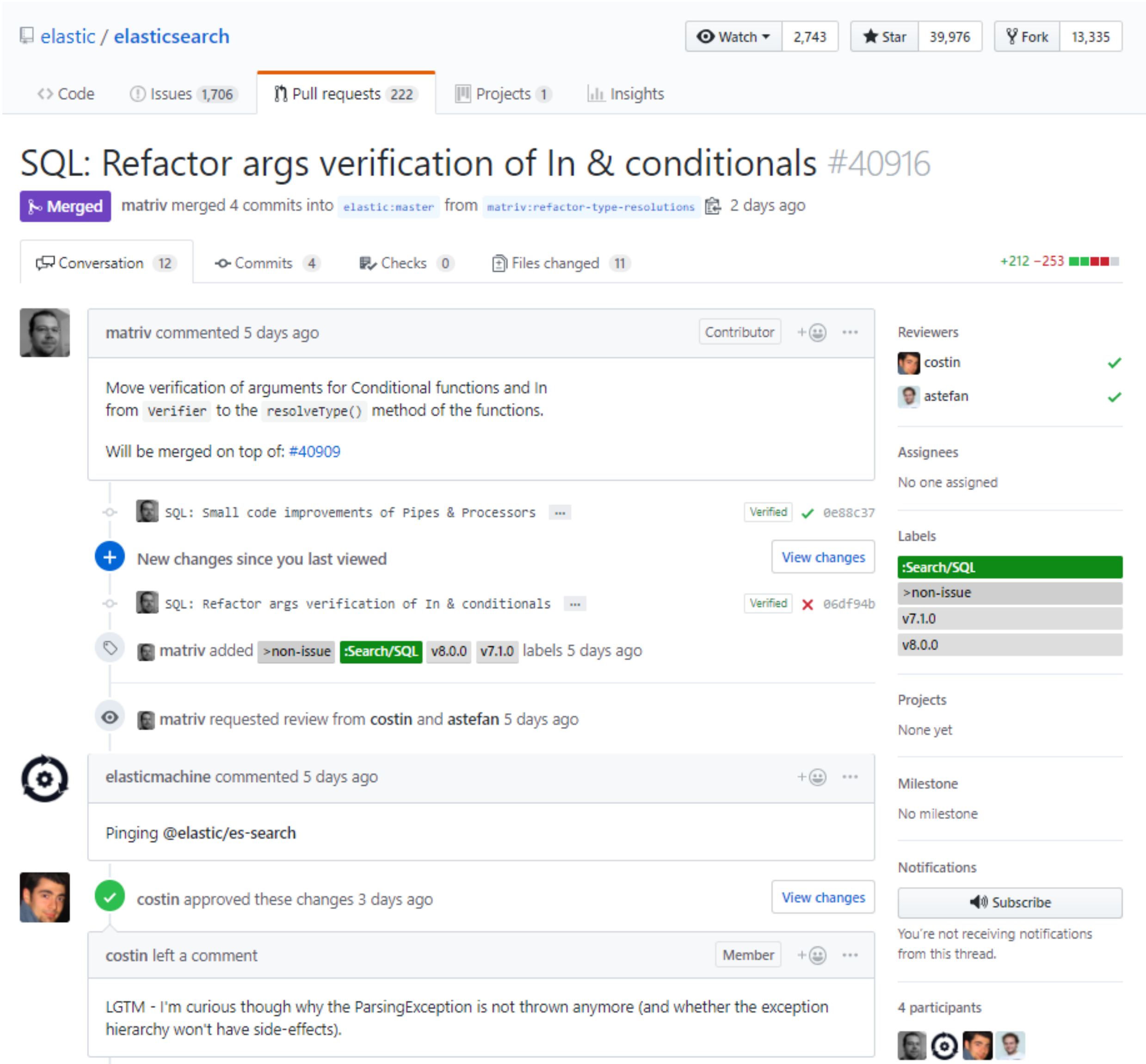
Figure 3.4A. Crowd collaboration on GitHub – example of Elastics (content)



*Source:[https://github.com/elastic/elasticsearch/pull/40916/commits/06df94b1247ce186af04f3538 46202843f7d8700](https://github.com/elastic/elasticsearch/pull/40916/commits/06df94b1247ce186af04f353846202843f7d8700)

* Green highlights of the source code are made by the contributor

Figure 3.4B. Crowd collaboration on GitHub – example of Elastic (communication)



Source: https://github.com/elastic/elasticsearch/pull/40916

Figure 3.5. Distribution of ventures' likelihood of crowd collaboration before and after matching

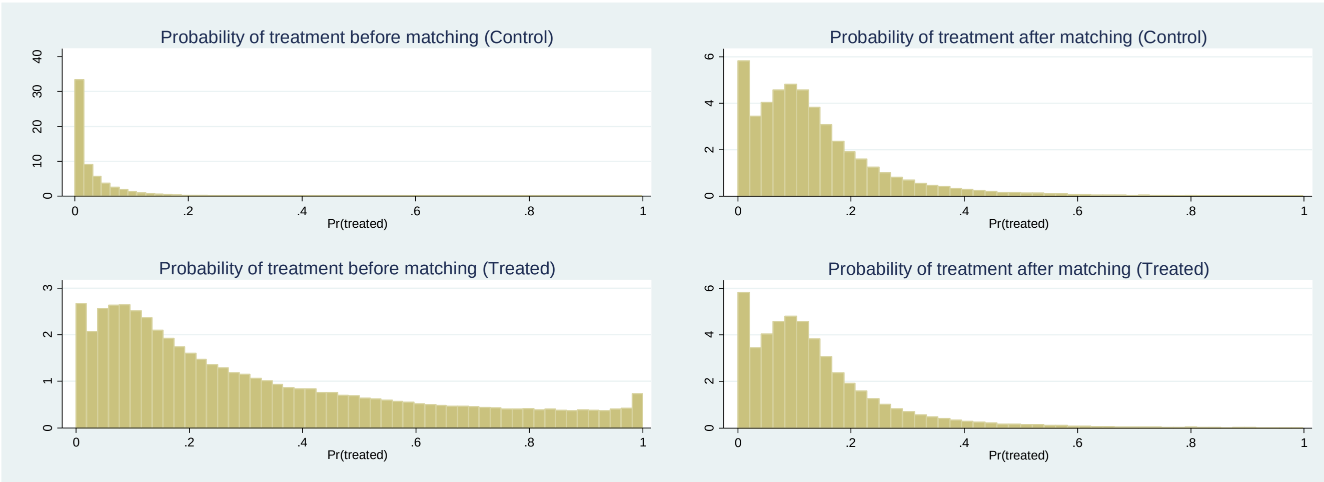


Figure 3.6. Sample ventures' frequency of crowd collaboration

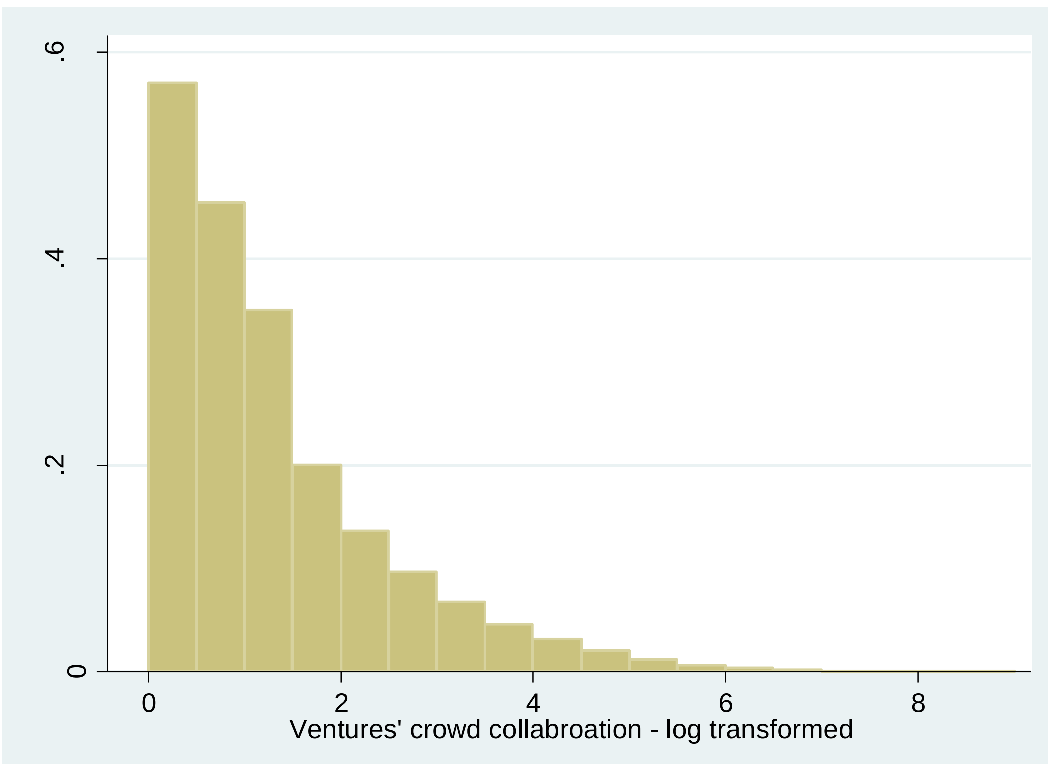
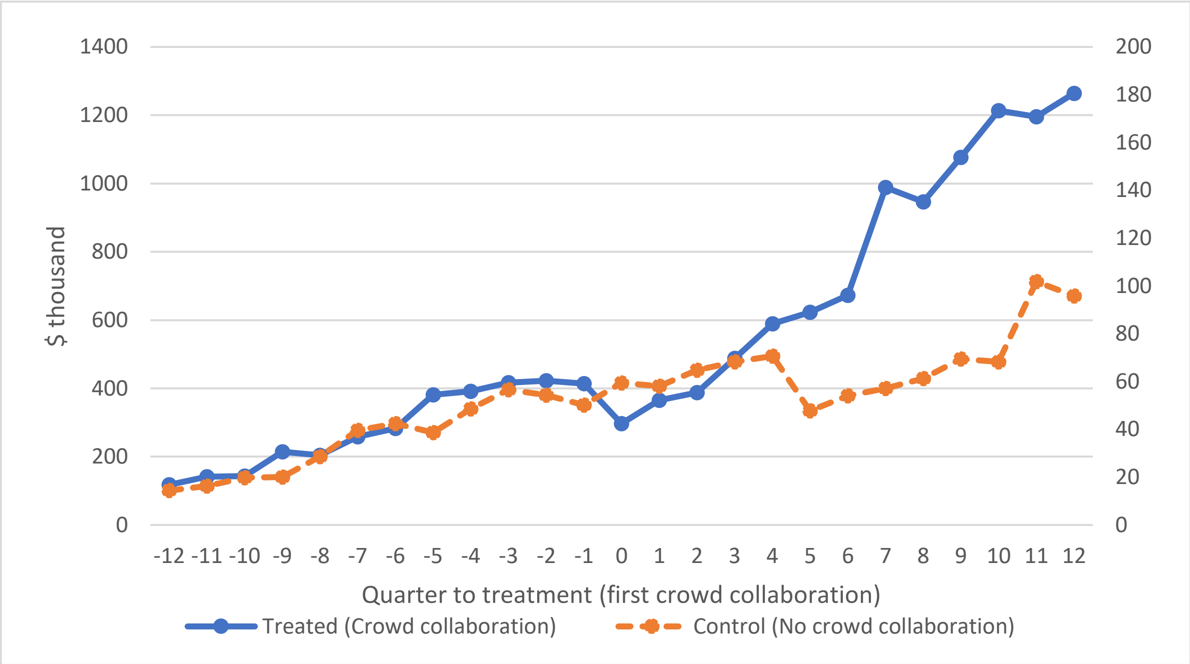


Figure 3.7. Average venture capital investment to open source based ventures pre-post crowd collaboration



References

Adner, R. 2017. Ecosystem as structure: an actionable construct for strategy. ***Journal of Management***, 43(1): 39-58.

Adner, R., & Kapoor, R. 2010. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. ***Strategic management journal***, 31(3): 306-333.

Adner, R., & Kapoor, R. 2016a. Innovation Ecosystems and the Pace of Substitution: Re-examining Technology S-curves. ***Strategic management journal***, 37(4): 625 - 648.

Adner, R., & Kapoor, R. 2016b. Innovation ecosystems and the pace of substitution: Re-examining technology S-curves. ***Strategic Management Journal***, 37(4): 625-648.

Afuah, A. 2001. Dynamic boundaries of the firm: are firms better off being vertically integrated in the face of a technological change? ***Academy of Management Journal***, 44(6): 1211-1228.

Afuah, A. 2013. Are network effects really all about size? the role of structure and conduct. ***Strategic Management Journal***, 34(3): 257-273.

Afuah, A., & Tucci, C. L. 2012. Crowdsourcing as a solution to distant search. ***Academy of management review***, 37(3): 355-375.

Afuah, A., & Tucci, C. L. 2013. Value capture and crowdsourcing. ***Academy of management review***, 38(3): 457-460.

Ahuja, G. 2000. Collaboration networks, structural holes, and innovation: A longitudinal study. ***Administrative science quarterly***, 45(3): 425-455.

Akerlof, G. A. 1970. The market for" lemons": Quality uncertainty and the market mechanism. ***The Quarterly Journal of Economics***: 488-500.

Alexy, O., George, G., & Salter, A. J. 2013a. Cui bono? The selective revealing of knowledge and its implications for innovative activity. ***Academy of Management Review***, 38(2): 270-291.

Alexy, O., Henkel, J., & Wallin, M. W. 2013b. From closed to open: Job role changes, individual predispositions, and the adoption of commercial open source software development. ***Research Policy***, 42(8): 1325-1340.

Alexy, O., & Reitzig, M. 2013. Private–collective innovation, competition, and firms' counterintuitive appropriation strategies. ***Research Policy***, 42(4): 895-913.

Alexy, O., West, J., Klapper, H., & Reitzig, M. 2018. Surrendering control to gain advantage: Reconciling openness and the resource-based view of the firm. ***Strategic Management Journal***, 39(6): 1704-1727.

Almirall, E., & Casadesus-Masanell, R. 2010. Open versus closed innovation: a model of discovery and divergence. ***Academy of management review***, 35(1): 27-47.

Anand, B. N., & Khanna, T. 2000. Do firms learn to create value? The case of alliances. ***Strategic management journal***, 21(3): 295-315.

Anderson, P., & Tushman, M. L. 1990. Technological discontinuities and dominant designs: A cyclical model of technological change. ***Administrative science quarterly***: 604-633.

Angrist, J. D., & Pischke, J.-S. 2008. ***Mostly harmless econometrics: An empiricist's companion***: Princeton university press.

Armstrong, M. 2006. Competition in two-sided markets. ***RAND Journal of Economics***, 00(3): 668-691.

Armstrong, M., & Wright, J. 2007. Two-Sided Markets, Competetive Bottlenecks and Exclusive Contracts: Research Article. ***Economic Theory***, 32(November): 353-380.

Asay, M. 2016. Who's No. 1 in open source? Microsoft! ***InFoWorld***.

Athey, S., Calvano, E., & Gans, J. S. 2011. The Impact of the Internet on Advertising Markets for News Media. ***Management Science***: 1-45.

Bagozzi, R. P., & Dholakia, U. M. 2006. Open source software user communities: A study of participation in Linux user groups. ***Management science***, 52(7): 1099-1115.

Baldwin, C., & von Hippel, E. 2011. Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. ***Organization Science***, 22(6): 1399-1417.

Baldwin, C. Y., & Clark, K. B. 2006. The architecture of participation: Does code architecture mitigate free riding in the open source development model? ***Management Science***, 52(7): 1116-1127.

Barney, J. 1991. Firm resources and sustained competitive advantage. ***Journal of management***, 17(1): 99-120.

Barney, J. B. 1986. Strategic factor markets: expectations, luck, and business strategy. ***Management science***, 32(10): 1231-1241.

Barney, J. B., Busenitz, L. W., Fiet, J. O., & Moesel, D. D. 1996. New venture teams' assessment of learning assistance from venture capital firms. ***Journal of Business Venturing***, 11(4): 257-272.

Baum, J. R., & Locke, E. A. 2004. The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. ***Journal of applied psychology***, 89(4): 587-598.

Bayus, B. L. 2013. Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. ***Management science***, 59(1): 226-244.

Belenzon, S., & Schankerman, M. 2015. Motivation and sorting of human capital in open innovation. ***Strategic Management Journal***, 36(6): 795-820.

Bianchi, A. J., Kang, S. M., & Stewart, D. 2012. The organizational selection of status characteristics: Status evaluations in an open source community. ***Organization Science***, 23(2): 341-354.

BlackDuck. 2015. 78% of Companies Run on Open Source Yet Lack Formal Policies.

Bloodgood, J. 2013. Crowdsourcing: Useful for problem solving, but what about value capture? ***Academy of management review***, 38(3): 455-457.

Bonaccorsi, A., Giannangeli, S., & Rossi, C. 2006. Entry strategies under competing standards: Hybrid business models in the open source software industry. ***Management Science***, 52(7): 1085-1098.

Bonaccorsi, A., & Rossi, C. 2003. Why open source software can succeed. ***Research policy***, 32(7): 1243-1258.

Boudreau, K. 2010. Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. ***Management Science***, 56(10): 1849-1872.

Boudreau, K. J. 2012. Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. ***Organization Science***, 23(5): 1409-1427.

Boudreau, K. J., & Jeppesen, L. B. 2014. Unpaid crowd complementors: The platform network effect mirage. ***Strategic management journal***.

Boudreau, K. J., & Jeppesen, L. B. 2015. Unpaid crowd complementors: The platform network effect mirage. ***Strategic Management Journal***, 36(12): 1761-1777.

Boudreau, K. J., & Lakhani, K. R. 2015. "Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. ***Research Policy***, 44(1): 4-19.

Bresnahan, T., Orsini, J., & Yin, P.-L. 2015. Demand heterogeneity, inframarginal multihoming , and platform market stability: mobile apps. ***Unpublished***.

Cabral, L. M. B. 1990. On the adoption of innovations with 'network' externalities. ***Mathematical Social Sciences***, 19(3): 299-308.

Caillaud, B., & Jullien, B. 2003. Chicken & egg: competition among intermediation service providers. ***RAND Journal of Economics***, 34(2): 309-328.

Carrillo, J. D., & Tan, G. 2006. Platform Competition: The Role of Multi-homing and Complementors.

Casadesus-Masanell, R., & Zhu, F. 2013. Business model innovation and competitive imitation: The case of sponsor-based business models. ***Strategic management journal***, 34(4): 464-482.

Cennamo, C., Ozalp, H., & Kretschmer, T. 2018. Platform Architecture and Quality Tradeoffs of Multihoming Complements. ***Information Systems Research***, In press.

Cennamo, C., & Santalo, J. 2013. Platform competition: Strategic trade-offs in platform markets. ***Strategic Management Journal***, 34(11): 1331-1350.

Chatterjee, S., & Wernerfelt, B. 1991. The link between resources and type of diversification: Theory and evidence. ***Strategic management journal***, 12(1): 33-48.

Chatterji, A. K., & Fabrizio, K. R. 2014. Using users: When does external knowledge enhance corporate product innovation? ***Strategic Management Journal***, 35(10): 1427-1445.

Chesbrough, H. W. 2003. ***Open innovation: The new imperative for creating and profiting from technology***: Harvard Business Press.

Chesbrough, H. W. 2006. The era of open innovation. ***Managing innovation and change***, 127(3): 34-41.

Choi. 2010. TYING IN TWO-SIDED MARKETS WITH MULTI-HOMING. ***Source: The Journal of Industrial Economics THE JOURNAL OF INDUSTRIAL ECONOMICS***, LXV(4): 872-886.

Christensen, J. F., Olesen, M. H., & Kjær, J. S. 2005. The industrial dynamics of Open Innovation—Evidence from the transformation of consumer electronics. ***Research Policy***, 34(10): 1533-1549.

Clarysse, B., Wright, M., Bruneel, J., & Mahajan, A. 2014. Creating value in ecosystems: Crossing the chasm between knowledge and business ecosystems. ***Research Policy***, 43(7): 1164-1176.

Coase, R. H. 1960. The Problem of Social Cost. ***Journal of law and economics***, 3(ArticleType: research-article / Full publication date: Oct., 1960 / Copyright © 1960 The University of Chicago): 1-44.

Cohen, W. M., & Levinthal, D. A. 1990. Absorptive capacity: a new perspective on learning and innovation. ***Administrative science quarterly***, 35(1).

Cohen, W. M., Nelson, R. R., & Walsh, J. P. 2000. Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not): National Bureau of Economic Research.

Colombo, M. G., & Grilli, L. 2010. On growth drivers of high-tech start-ups: Exploring the role of founders' human capital and venture capital. ***Journal of Business Venturing***, 25(6): 610-626.

Colombo, M. G., Piva, E., & Rossi-Lamastra, C. 2014. Open innovation and within-industry diversification in small and medium enterprises: The case of open source software firms. ***Research Policy***, 43(5): 891-902.

Comino, S., Manenti, F. M., & Parisi, M. L. 2007. From planning to mature: On the success of open source projects. ***Research Policy***, 36(10): 1575-1586.

Corts, K. S., & Lederman, M. 2008. Software exclusivity and the scope of indirect network effects in the U.S. home video game market ☆. ***International Journal of Industrial Organization***, 27: 121-136.
Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. 2012. ***Social coding in GitHub: transparency and collaboration in an open software repository***. Paper presented at the Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work.
Dahlander, L., & Frederiksen, L. 2012. The core and cosmopolitans: A relational view of innovation in user communities. ***Organization science***, 23(4): 988-1007.
Dahlander, L., & Magnusson, M. G. 2005. Relationships between open source software companies and communities: Observations from Nordic firms. ***Research policy***, 34(4): 481-493.
Dahlander, L., & Piezunka, H. 2014. Open to suggestions: How organizations elicit suggestions through proactive and reactive attention. ***Research Policy***, 43(5): 812-827.
Davila, A., Foster, G., & Gupta, M. 2003. Venture capital financing and the growth of startup firms. ***Journal of Business Venturing***, 18(6): 689-708.
De Clercq, D., & Sapienza, H. J. 2006. Effects of relational capital and commitment on venture capitalists' perception of portfolio company performance. ***Journal of Business Venturing***, 21(3): 326-347.
Dierickx, I., & Cool, K. 1989. Asset stock accumulation and sustainability of competitive advantage. ***Management science***, 35(12): 1504-1511.
Dokko, G., & Rosenkopf, L. 2010. Social capital for hire? Mobility of technical professionals and firm influence in wireless standards committees. ***Organization Science***, 21(3): 677-695.
Donald, S. G., & Lang, K. 2007. Inference with difference-in-differences and other panel data. ***The review of economics and statistics***, 89(2): 221-233.
Dosi, G. 1982. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. ***Research Policy***, 11(3): 147-162.
Doz, Y. L. 1996. The evolution of cooperation in strategic alliances: Initial conditions or learning processes? ***Strategic management journal***, 17(S1): 55-83.
Dua, A. 2017. LINUX IS RUNNING ON ALMOST ALL OF THE SUPERCOMPUTERS.
Ebner, W., Leimeister, J. M., & Krcmar, H. 2009. Community engineering for innovations: the ideas competition as a method to nurture a virtual community for innovations. ***R&d Management***, 39(4): 342-356.
Economides, N., & Katsamakas, E. 2006. Two-sided competition of proprietary vs. open source technology platforms and the implications for the software industry. ***Management Science***, 52(7): 1057-1071.
Eisenhardt, K. M., & Martin, J. A. 2000. Dynamic capabilities: what are they? ***Strategic management journal***, 21(10-11): 1105-1121.
Eisenmann, T., Parker, G., & Van Alstyne, M. 2011. Platform envelopment. ***Strategic management journal***, 32(12): 1270-1285.
Evans, D. S. 2003. Some empirical aspects of multi-sided platform industries. ***Review of Network Economics***, 2(3).
Farrell, J., & Klemperer, P. 2007. Chapter 31 Coordination and Lock-In: Competition with Switching Costs and Network Effects, Vol. 3: 1967-2072.
Feldman, E. R. 2015. Corporate spinoffs and analysts' coverage decisions: The implications for diversified firms. ***Strategic Management Journal***.
Felin, T., & Zenger, T. R. 2014. Closed or open innovation? Problem solving and the governance choice. ***Research Policy***, 43(5): 914-925.

Fitzgerald, B. 2006. The transformation of open source software. ***Mis Quarterly***: 587-598.
Fleming, L., & Waguespack, D. M. 2007. Brokerage, boundary spanning, and leadership in open innovation communities. ***Organization science***, 18(2): 165-180.
Fosfuri, A., Giarratana, M. S., & Luzzi, A. 2008. The penguin has entered the building: The commercialization of open source software products. ***Organization science***, 19(2): 292-305.
Foss, N. J., Frederiksen, L., & Rullani, F. 2016. Problem-formulation and problem-solving in self-organized communities: How modes of communication shape project behaviors in the free open-source software community. ***Strategic management journal***, 37(13): 2589-2610.
Franke, N., & Von Hippel, E. 2003. Satisfying heterogeneous user needs via innovation toolkits: the case of Apache security software. ***Research policy***, 32(7): 1199-1215.
Frederiksen, L., & Rullani, F. 2016. Problem-formulation and problem-solving in self-organized communities: How modes of communication shape project behaviors in the free open-source software community. ***Strategic Management Journal***, 37(13): 2589-2610.
Gans, J. S., & Stern, S. 2003. The product market and the market for “ideas”: commercialization strategies for technology entrepreneurs. ***Research policy***, 32(2): 333-350.
Garriga, H., Von Krogh, G., & Spaeth, S. 2013. How constraints and knowledge impact open innovation. ***Strategic Management Journal***, 34(9): 1134-1144.
Gawer, A., & Cusumano, M. A. 2002. ***Platform leadership: How Intel, Microsoft, and Cisco drive industry innovation***: Harvard Business School Press Boston.
Gawer, A., & Henderson, R. 2007. Platform owner entry and innovation in complementary markets: Evidence from Intel. ***Journal of Economics & Management Strategy***, 16(1): 1-34.
Gompers, P., & Lerner, J. 2001. The venture capital revolution. ***Journal of Economic Perspectives***: 145-168.
Gousios, G. 2013. ***The GHTorent dataset and tool suite***. Paper presented at the Proceedings of the 10th working conference on mining software repositories.
Grant, R. M. 1996. Toward a knowledge-based theory of the firm. ***Strategic management journal***, 17(S2): 109-122.
Grigorik, I. 2012. The github archive. ***URL: https://githubarchive. org***.
Gruber, M., & Henkel, J. 2006. New ventures based on open innovation–an empirical analysis of start-up firms in embedded Linux. ***International Journal of Technology Management***, 33(4): 356-372.
Gu, R., Oh, L. B., & Wang, K. 2016. Multi-homing on SNSs: The role of optimum stimulation level and perceived complementarity in need gratification. ***Information and Management***, 53(6): 752-766.
Gulati, R. 1995. Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. ***Academy of Management Journal***, 38(1): 85-112.
Gulati, R. 1998. Alliances and networks. ***Strategic management journal***, 19(4): 293-317.
Gulati, R., & Singh, H. 1998. The architecture of cooperation: Managing coordination costs and appropriation concerns in strategic alliances. ***Administrative science quarterly***: 781-814.
Haefliger, S., Von Krogh, G., & Spaeth, S. 2008. Code reuse in open source software. ***Management Science***, 54(1): 180-193.
Hagiu, A. 2006. Pricing and commitment by two-sided platforms. ***RAND Journal of Economics Journal of Economics***, 12(3): 720-737.
Hagiu, A. 2009. Two-sided platforms: Product variety and pricing structures. ***Journal of Economics and Management Strategy***, 18(4): 1011-1043.

Hagiu, C., Wright, J., Andrei Hagiu, B., Hagiu Julian Wright, A., & Hagiu, A. 2018. Multi-Sided Platforms Multi-Sided Platforms Multi-Sided Platforms *. ***International Journal of Industrial Organization***.

Hargadon, A., & Sutton, R. I. 1997. Technology brokering and innovation in a product development firm. ***Administrative science quarterly***: 716-749.

Henkel, J. 2006. Selective revealing in open innovation processes: The case of embedded Linux. ***Research policy***, 35(7): 953-969.

Henkel, J., Schöberl, S., & Alexy, O. 2014. The emergence of openness: How and why firms adopt selective revealing in open innovation. ***Research Policy***, 43(5): 879-890.

Hertel, G., Niedner, S., & Herrmann, S. 2003. Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. ***Research policy***, 32(7): 1159-1177.

Hippel, E. v., & Krogh, G. v. 2003. Open source software and the "private-collective" innovation model: Issues for organization science. ***Organization science***, 14(2): 209-223.

Holland, P. W. 1986. Statistics and Causal Inference. ***Journal of the American Statistical Association***, 81(396): 945-960.

Howe, J. 2008. ***Crowdsourcing: How the power of the crowd is driving the future of business***: Random House.

Hyrynsalmi, S., Suominen, A., & Mäntymäki, M. 2015. The influence of developer multi-homing on competition between software ecosystems. ***The Journal of Systems and Software***, 111: 119-127.

Jacks, J. 2018. The $100B+ Commercial OSS Company Index (COSSCI).

Jeppesen, L. B., & Frederiksen, L. 2006. Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. ***Organization science***, 17(1): 45-63.

Jeppesen, L. B., & Lakhani, K. R. 2010. Marginality and problem-solving effectiveness in broadcast search. ***Organization Science***, 21(5): 1016-1033.

Kalnins, A. 2018. Multicollinearity: How common factors cause Type 1 errors in multivariate regression. ***Strategic Management Journal***, 39(8): 2362-2385.

Kapoor, R., & Agarwal, S. 2017. Sustaining superior performance in business ecosystems: Evidence from application software developers in the iOS and Android smartphone ecosystems. ***Organization Science***, 28(3): 531-551.

Kapoor, R., & Furr, N. R. 2015. Complementarities and competition: Unpacking the drivers of entrants' technology choices in the solar photovoltaic industry. ***Strategic Management Journal***, 36(3): 416-436.

Kapoor, R., & Lee, J. M. 2013. Coordinating and competing in ecosystems: How organizational forms shape new technology investments. ***Strategic management journal***, 34(3): 274-296.

Katila, R., & Ahuja, G. 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. ***Academy of Management Journal***, 45(6): 1183-1194.

Katz, M. L., & Shapiro, C. 1986. Technology Adoption in the Presence of Network Externalities. ***Source Journal of Political Economy***, 94(4): 822-841.

Katz, M. L., & Shapiro, C. 1994. Systems Competition and Network Effects. ***Journal of Economic Perspectives—Volume***, 8(2—Spring): 93-115.

Kleinbaum, D., Kupper, L., Muller, K., & Nizam, A. 1988. The correlation coefficient and straight-line regression analysis. ***Applied Regression Analysis and Other Multivariable Methods. Belmont, CA: Wadsworth Publishing Company***: 80-95.

Kogut, B., & Metiu, A. 2001. Open-source software development and distributed innovation. ***Oxford review of economic policy***, 17(2): 248-264.
Kogut, B., & Zander, U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. ***Organization Science***, 3(3): 383-397.
Krishnamurthy, S., Ou, S., & Tripathi, A. K. 2014. Acceptance of monetary rewards in open source software development. ***Research Policy***, 43(4): 632-644.
Krishnan, H. A., Miller, A., & Judge, W. Q. 1997. Diversification and top management team complementarity: Is performance improved by merging similar or dissimilar teams? ***Strategic management journal***, 18(5): 361-374.
Kuk, G. 2006. Strategic interaction and knowledge sharing in the KDE developer mailing list. ***Management Science***, 52(7): 1031-1042.
Lakhani, K. R., & Von Hippel, E. 2003. How open source software works:"free" user-to-user assistance. ***Research policy***, 32(6): 923-943.
Landsman, V., & Stremersch, S. 2011. Multihoming in Two-Sided Markets: An Empirical Inquiry in the Video Game Console Industry. ***Journal of Marketing***, 75(6): 39-54.
Lee, E., Lee, J., & Lee, J. 2006. Reconsideration of the Winner-Take-All Hypothesis: Complex Networks and Local Bias. ***Management Science***, 52(12).
Lee, G. K., & Cole, R. E. 2003. From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. ***Organization science***, 14(6): 633-649.
Leiponen, A. E. 2008. Competing through cooperation: The organization of standard setting in wireless telecommunications. ***Management Science***, 54(11): 1904-1919.
Lerner, & Tirole. SOME SIMPLE ECONOMICS OF OPEN SOURCE*.
Lerner, J., & Tirole, J. 2005a. The economics of technology sharing: Open source and beyond. ***Journal of Economic Perspectives***, 19(2): 99-120.
Lerner, J., & Tirole, J. 2005b. The scope of open source licensing. ***Journal of Law, Economics, and Organization***, 21(1): 20-56.
Levine, S. S., & Prietula, M. J. 2013. Open collaboration for innovation: principles and performance. ***Organization Science***, 25(5): 1414-1433.
Levinthal, D. A. 1997. Adaptation on rugged landscapes. ***Management science***, 43(7): 934-950.
Lichtenthaler, U. 2011. Open innovation: Past research, current debates, and future directions. ***The Academy of Management Perspectives***, 25(1): 75-93.
Liebowitz, S. J., & Margolis, S. E. 1994. Network externality: An uncommon tragedy. ***Journal of economic perspectives***, 8(2): 133-150.
Loyalka, P., Liu, O. L., Li, G., Chirikov, I., Kardanova, E., Gu, L., Ling, G., Yu, N., Guo, F., & Ma, L. 2019. Computer science skills across China, India, Russia, and the United States. ***Proceedings of the National Academy of Sciences***: 201814646.
MacCormack, A., Rusnak, J., & Baldwin, C. Y. 2006. Exploring the structure of complex software designs: An empirical study of open source and proprietary code. ***Management Science***, 52(7): 1015-1030.
March, J. G. 1991. Exploration and exploitation in organizational learning. ***Organization Science***, 2(1): 71-87.
Markides, C. C., & Williamson, P. J. 1994. Related diversification, core competences and corporate performance. ***Strategic management journal***, 15(S2): 149-165.
Massa, L., Tucci, C. L., & Afuah, A. 2017. A critical assessment of business model research. ***Academy of Management Annals***, 11(1): 73-104.

McIntyre, D. P., & Srinivasan, A. 2017. Networks, platforms, and strategy: Emerging views and next steps. ***Strategic Management Journal***, 38(1): 141-160.
Mckinsey. 2013. Open data: Unlocking innovation and performance with liquid information.
Miller, D. J. 2006. Technological diversity, related diversification, and firm performance. ***Strategic Management Journal***, 27(7): 601-619.
Montgomery, C. A., & Wernerfelt, B. 1988. Diversification, Ricardian rents, and Tobin's q. ***The Rand journal of economics***: 623-632.
Mowery, D. C., Oxley, J. E., & Silverman, B. S. 1996. Strategic alliances and interfirm knowledge transfer. ***Strategic management journal***, 17(S2): 77-91.
Myers, A. 2018. 2018: The Biggest Year for Open Source Software Ever! ***Medium***.
Nagaraj, A., & Piezunka, H. 2017. Motivating the Established, Deterring the New—The Divergent Effect of Platform Competition on Member Contributions in Digital Mapping Communities.
Nagle, F. 2018. Open source software and firm productivity. ***Management Science***.
Nelson, R. R., & Winter, S. G. 1982. ***An evolutionary theory of economic change***: Harvard University Press.
Nesbitt, A., & Pompilio, M. 2016. Libraries. io.
Nickerson, J. A., & Zenger, T. R. 2004. A knowledge-based theory of the firm—The problem-solving perspective. ***Organization Science***, 15(6): 617-632.
O'Mahony, S., & Ferraro, F. 2007. The emergence of governance in an open source community. ***Academy of Management Journal***, 50(5): 1079-1106.
O'Mahony, S. 2003. Guarding the commons: how community managed software projects protect their work. ***Research Policy***, 32(7): 1179-1198.
Octoverse. 2018. The State of the Octoverse 2017.
Oh, W., & Jeon, S. 2007. Membership herding and network stability in the open source community: The Ising perspective. ***Management science***, 53(7): 1086-1101.
Oxley, J. E. 1997. Appropriability hazards and governance in strategic alliances: A transaction cost approach. ***Journal of Law, Economics, and Organization***, 13(2): 387-409.
Parker, G., & Alstyne, M. V. 2017. Innovation, Openness, and Platform Control. ***Management Science***.
Parker, G. G., & Alstyne, M. W. V. 2005. Two-Sided Network Effects: A Theory of Information Product Design. ***Source: Management Science***, 51(10): 1494-1504.
Penrose, E. T. 1959. ***The Theory of the Growth of the Firm***: Oxford university press.
Peteraf, M. A. 1993. The cornerstones of competitive advantage: a resource-based view. ***Strategic management journal***, 14(3): 179-191.
Piezunka, H., & Dahlander, L. 2015. Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. ***Academy of Management Journal***, 58(3): 856-880.
Pisano, G. P. 1990. The R&D boundaries of the firm: an empirical analysis. ***Administrative science quarterly***: 153-176.
Rangan, S., & Adner, R. 2001. Profits and the Internet: Seven misconceptions. ***MIT Sloan Management Review***, 42(4): 44.
Ranganathan, R., Ghosh, A., & Rosenkopf, L. 2018. Competition-Cooperation interplay during Multi-Firm technology coordination. ***Strategic Management Journal***.
Ranganathan, R., & Rosenkopf, L. 2014. Do ties really bind? The effect of knowledge and commercialization networks on opposition to standards. ***Academy of Management Journal***, 57(2): 515-540.

Rasch, A. 2007. Platform competition with partial multihoming under differentiation: A note. ***Economics Bulletin***, 12(7): 1-8.
Raymond, E. S. 2001. The cathedral and the bazaar-musings on Linux and open source by an accidental revoltionary (rev. ed.): O'reilly Beijing.
Roberts, J. A., Hann, I.-H., & Slaughter, S. A. 2006. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. ***Management science***, 52(7): 984-999.
Rochet, J.-C., & Tirole, J. 2003. Two-Sided Markets. ***Journal of the European Economic Association***: 990-1029.
Rochet, J. C., & Tirole, J. 2006. Two-sided markets: a progress report. ***The RAND journal of economics***, 37(3): 645-667.
Rogers, E. M. 2010. ***Diffusion of innovations***: Simon and Schuster.
Rosen, L. 2004. ***Open source licensing: Software freedom and intellectual property law***: Prentice Hall PTR.
Rosenkopf, L., Metiu, A., & George, V. P. 2001. From the bottom up? Technical committee activity and alliance formation. ***Administrative Science Quarterly***, 46(4): 748-772.
Rosenkopf, L., & Tushman, M. L. 1998. The coevolution of community networks and technology: Lessons from the flight simulation industry. ***Industrial and Corporate Change***, 7(2): 311-346.
Rowley, J. D. 2017. Open Source Software Is Big Business With Big Funding.
Rumelt, R. P. 1984. Competitive strategic management. ***Toward a Strategic Theory of the Firm***: 556-570.
Rysman, M., & Simcoe, T. 2008. Patents and the performance of voluntary standard-setting organizations. ***Management science***, 54(11): 1920-1934.
Sakhartov, A. V. 2018. Stock market undervaluation of resource redeployability. ***Strategic Management Journal***, 39(4): 1059-1082.
Schilling, M. 2002a. Technology Success and Failure in Winner Take All Markets: The Impact of Learning Orientation, Timing and Network Externalities. ***Academy of Management Journal***, 45(2): 387-398.
Schilling, M. A. 2002b. Technology success and failure in winner-take-all markets: The impact of learning orientation, timing, and network externalities. ***Academy of Management Journal***, 45(2): 387-398.
Seamans, R., & Zhu, F. 2014. Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers. ***Management Science***, 60(2): 476-493.
Shah, S. K. 2006. Motivation, governance, and the viability of hybrid forms in open source software development. ***Management Science***, 52(7): 1000-1014.
Shankar, V., & Bayus, B. L. 2003. Network effects and competition: An empirical analysis of the home video game industry. ***Strategic Management Journal***, 24(4): 375-384.
Shapiro, C., & Varian, H. R. 1998. ***Information rules: a strategic guide to the network economy***: Harvard Business Press.
Silver, A. 2018. Microsoft's purchase of GitHub leaves some scientists uneasy. ***Nature***, 558(7710): 353-353.
Solomon, G. 2018. Why the Elastic IPO is So Important, ***Goinglongblog***.
Spence, M. 1974. Competitive and optimal responses to signals: An analysis of efficiency and distribution. ***Journal of Economic theory***, 7(3): 296-332.
StackExchange. 2015. What are dependencies?

Stam, W. 2009. When does community participation enhance the performance of open source software companies? ***Research Policy***, 38(8): 1288-1299.
Stiglitz, J. E. 1975. The theory of" screening," education, and the distribution of income. ***The American economic review***, 65(3): 283-300.
Stuart, T. E. 2000. Interorganizational alliances and the performance of firms: A study of growth and innovation rates in a high-technology industry. ***Strategic management journal***, 21(8): 791-811.
Suarez, F. F. 2005. Network effects revisited: The role of strong ties in technology selection. ***Academy of Management Journal***, 48(4): 710-720.
Suarez, F. F., & Utterback, J. M. 1995. Dominant designs and the survival of firms. ***Strategic management journal***, 16(6): 415-430.
Techfae. 2016. LINUX IS RUNNING ON ALMOST ALL OF THE SUPERCOMPUTERS.
Techworm. 2016. What is the difference between Linux and UNIX operating systems?
Teece, D. J. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. ***Research Policy***, 15(6): 285-305.
Teece, D. J. 2007. Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. ***Strategic management journal***, 28(13): 1319-1350.
Toh, P. K., & Miller, C. D. 2017. Pawn to Save a Chariot, or Drawbridge Into the Fort? Firms' Disclosure During Standard Setting and Complementary Technologies Within Ecosystems. ***Strategic Management Journal***.
Tucci, C. L., Afuah, A., & Viscusi, G. 2018. ***Creating and capturing value through crowdsourcing***: Oxford University Press.
Venkatraman, N., & Lee, C.-H. 2004. Preferential Linkage and Network Evolution: A Conceptual Model and Empirical Test in the U.S. Video Game Sector. ***Academy of Management Journal***, 47(6): 876-892.
Von Hippel, E. 1986. Lead users: a source of novel product concepts. ***Management science***, 32(7): 791-805.
Von Krogh, G., Spaeth, S., & Lakhani, K. R. 2003. Community, joining, and specialization in open source software innovation: A case study. ***Research Policy***, 32(7): 1217-1241.
Von Krogh, G., & Von Hippel, E. 2003. Special issue on open source software development: Elsevier.
Von Krogh, G., & Von Hippel, E. 2006. The promise of research on open source software. ***Management science***, 52(7): 975-983.
Waguespack, D. M., & Fleming, L. 2009. Scanning the commons? Evidence on the benefits to startups participating in open standards development. ***Management Science***, 55(2): 210-223.
Wen, W., Ceccagnoli, M., & Forman, C. 2015. Opening up intellectual property strategy: Implications for open source software entry by start-up firms. ***Management Science***, 62(9): 2668-2691.
Wen, W., Forman, C., & Graham, S. J. 2013. Research note—The impact of intellectual property rights enforcement on open source software project success. ***Information Systems Research***, 24(4): 1131-1146.
Wernerfelt, B. 1984. A resource-based view of the firm. ***Strategic management journal***, 5(2): 171-180.
West, J. 2003. How open is open enough?: Melding proprietary and open source platform strategies. ***Research policy***, 32(7): 1259-1285.

West, J., & Gallagher, S. 2006. Challenges of open innovation: the paradox of firm investment in open-source software. ***R&d Management***, 36(3): 319-331.

Wikipedia. 2018a. Library (computing).

Wikipedia. 2018b. Open source model.

Wikipedia. 2018c. Package manager.

Wikipedia. 2018d. Version Control.

Williamson, O. E. 1979. Transaction-cost economics: the governance of contractual relations. ***Journal of law and economics***, 22(2): 233-261.

Williamson, O. E. 1985. ***The economic intstitutions of capitalism***: Simon and Schuster.

Williamson Oliver, E. 1985. The economic institutions of capitalism: Firms, markets, relational contracting. ***New York***.

Wired. 2016. The Problem With Putting All the World's Code in GitHub.

Wooldridge, J. 2012. ***Introductory econometrics: A modern approach***: Cengage Learning.

Wu, B. 2013. Opportunity costs, industry dynamics, and corporate diversification: Evidence from the cardiovascular medical device industry, 1976–2004. ***Strategic management journal***, 34(11): 1265-1287.

Zeitlyn, D. 2003. Gift economies in the development of open source software: anthropological reflections. ***Research policy***, 32(7): 1287-1291.

Zhou, Y. M. 2011. Synergy, coordination costs, and diversification choices. ***Strategic Management Journal***, 32(6): 624-639.

Zhu, F., & Iansiti, M. 2012. Entry into platform-based markets. ***Strategic management journal***, 33(1): 88-106.