INDIVIDUAL INTERACTION AND INNOVATION CAPABILITIES: EXPLORATION AND EXPLOITATION IN OPEN SOURCE SOFTWARE COMMUNITIES

Francesco Rullani
INO, Copenhagen Business School
fr.ino@cbs.dk

Lars Frederiksen
I&E Group, Imperial College Business School
l.frederiksen@imperial.ac.uk

Abstract:
This paper analyzes the micro-foundation for development of innovation capabilities in an open innovation context. To generate our results we employ an unbalanced panel of 2,598 Open Source Software developers working on projects hosted on the SourceForge online community over 28 months. We theorize how interaction in a self-organized community translates into individuals communicative behaviors. Next, we investigate empirically how individual interaction affects individual behaviors, and how the difference in individual behavior maps into community-level capabilities. We hypothesize that the type of media through which individuals interact influences the amount of resources accumulated for community exploration rather than exploitation. Our estimations show that the more community interaction is undertaken through verbal communication, the less the resources allocated to exploration and the more to exploitation, while communication based on exchange of artifacts e.g. patch files leads to the opposite results. Finally, we draw theoretical conclusions and offer managerial implications based on our study.

JEL - codes: O31, L86, M00
Individual Interaction and Innovation Capabilities

Exploration and Exploitation in Open Source Software Communities

DRUID Summer Conference 2010

London

Word count: main text 8.250
Introduction

“…there is a widespread sense of a gap between the rapid development of new organizational forms in practice and the capacity of existing perspectives to account for them in theory” (Child and McGrath, 2001, p. 1135)

To advance our understanding of how capabilities emerge and accumulate, we need to employ novel theoretical perspectives as well as new empirical domains. These ‘tools’ may enable us to progress towards identifying different significant mechanisms underlying capabilities development.

Scholars have argued - convincingly - that an organization and its capabilities essentially are created and accumulated by individuals integrating and recombining their knowledge bases through different forms of social interaction (Barnard, 1938; Nelson and Winter, 1982; Ouchi, 1990; Kogut and Zander, 1992). There are several ways such interaction may take place. Burns and Stalker (1961) argue that the organic organization as a ‘communities of interest’ (i.e. loosely coupled networks of employees) is better able to adapt to dynamic environments. Kellogg, Orlikowski and Yates (2006) extend Burns and Stalker’s idea to the concept of the post-bureaucratic organizational form, where decisions are based on dialogue and consensus, the organization is a network, and boundaries are permeable. A pioneering continuation of the increasingly ‘loose’ mode of organizing individual interaction is through a self-organizing structure (Kaufman, 1995). Yet, self-organizing structures are unexplored territory for research on capabilities, but in line with the intentions of the multifaceted research stream on capability development (Kogut and Zander 1992; Helfat and Peteraf, 2003). Despite the increasing number of instances of self-organization (e.g. Anderson, 1999), little is known about how capabilities are built up
through individual interaction in self-organizing environments, where individuals enjoy high degrees of freedom in choosing which tasks to undertake and in allocating their efforts, time and energy to different activities.

In this paper, we seek to advance our understanding of this topic, arguing that different types of interaction between individuals lead to the accumulation of different types of capabilities. Following the general distinction made by DeSantis and Monge (1999), we differentiate individual interactions as 1) what one does together with others (i.e. task) and 2) what one says to others (i.e. verbal communication). Organizational structures going in the direction of self-organization are often hailed as spurring and sustaining innovativeness (von Hippel and Von Krogh, 2003). We focus on the innovation capabilities of exploration and exploitation (March, 1991) and on their specific features in self-organizing settings compared to more traditional organizational environments (Gupta et al., 2006).

Our investigation implements an operationalization and empirical application of the conceptual model presented by Felin and Foss (2005) relating macro-aggregates observed at the level of the organization, such as capabilities, to micro-behaviors relating to the actions and interactions of individuals. This choice of conceptual framework is driven by the nature of the phenomenon we are investigating: self-organization as a bottom-up process in which macro-structures are simultaneously the context for individuals’ actions and determined by the actions and interactions of individuals (Giddens, 1984).

To generate our results we employed two Poisson estimations on a dataset consisting of

The findings support our claim that if a self-organizing community structures its interactions around the exchange of artifacts (in the OSS case, source code), the property that emerges from this interaction results in the allocation of *more* resources (i.e. a higher number of individual developers) for community-level capabilities related to exploitation and fewer resources for exploration. In contrast, when the interaction structure in the community relies more on verbal communication (e.g. online exchange of messages), we find that more resources are allocated to exploration and fewer are allocated to exploitation.

These findings yield insights for research on capabilities (Finkelstein et al., 2007). Our study contributes in three ways: first, by showing how the development of exploration and exploitation capabilities at the collective level of analysis is dependent on different types of individual action and social interaction. By doing so we complement Fang et al. (forthcoming 2010) study of the relationship between the capabilities of exploration and exploitation, and the network structure of the individuals’ relations. Besides, we extend Agterberg’s et al. (forthcoming 2010) analysis of how content and connections affect knowledge sharing in intra-organizational networks of practice. Here we underline the importance of the specific type (i.e. quantity and quality) of individual interactions required for capabilities to develop. Second, we develop arguments that open up the discussion on capability development and innovation in the OSS community, thus
shedding light on the features of a research setting emblematic of increasingly central knowledge-based communities and distributed innovation (David and Foray, 2003; Sproull et al., 2008; Dahlander et al., 2008; Jeppesen and Frederiksen, 2006). Third, in self-organized organizations: "... managers establish and modify the direction and the boundaries within which effective, improvised, self-organized solutions can evolve ... and tune the system by altering the constraints." (Anderson, 1999, p. 228). We provide managers with a vision of the tools that can be applied to shape the strategies of the self-organizing groups in which they are involved, by investigating the effects of the quantity and quality of interactions on individuals’ behaviors. This is a crucial undertaking because, as Agterberg et al. (forthcoming, p. 87) put it: "... how to balance between emergent self-organization and autonomy on the one hand ... and some degree of formal management influence or control on the other hand? ... Knowledge management literature ... cannot sufficiently explain yet how organizations deal with these conflicting demands..."

The paper is organized as follows: Section 2 provides the theoretical background to our study. Section 3 briefly explains the specificities of OSS communities and discusses how the theoretical model presented in Section 2 is adapted to fit the specificities of the research setting. In Section 4 we formulate four hypotheses regarding innovation capabilities. Section 5 describes the methodology and data used and Sections 6 and 7 respectively present and discuss the results. Section 8 offers some conclusions.

2. Theoretical background

We bring together two streams of literature: studies on capability development and
accumulation (Zollo and Winter, 2002; Teece et al., 1997; Helfat and Peteraf, 2003) and research on knowledge-based communities (Brown and Duguid, 1991; David and Foray, 2003; Cole and Lee, 2003). We bridge these literatures using a conceptual lens that is gaining increasing interest among management scholars: micro-foundations of organizational level constructs (Foss and Felin, 2005).

Traditionally, the notion of capabilities is attributed to the ‘service’ of individual managers in bundling resources within firms (Penrose, 1959), leading to heterogeneity among firms and thus potentially creation of competitive advantage (Makadok, 2001). An upsurge of research is moving beyond Penrose (1959) and Richardson (1972)¹ to tackle the question of how capabilities emerge, evolve, and erode. Researchers have identified important explanatory variables such as luck or superior information (Barney, 1986), deliberate collective learning processes (Zollo and Winter, 2002), lifecycles of capabilities (Helfat and Peteraf, 2003), and dynamic capabilities (Teece et al., 1997). Most of these accounts agree that “…capabilities are not only manifestations of observable corporate structures, but also reside in corporate cultures and network relations of employees. Yet, capabilities do not vest in a single individual, nor are they capable of being articulated by any individual since they are supra-individual…” (Collis, 1994, p. 145). Others offer an alternative view and focused in their analysis on the specific individual qualities of cognition (Tripsas and Gavetti, 2000), framing (Kaplan, 2008), and motivations (Gottschalg and Zollo, 2007) to explain the mechanisms of capability change over time. However, most of the studies concerned with capability accumulation analyze organizational settings with some degree of direction, pecuniary incentive, and formal hierarchy.
In the organizational learning literature, March (1991) launched the argument that an organization’s innovation capabilities include both explorative and exploitative abilities. Different types of learning and interaction underlie the origin and development of these dissimilar capabilities in the organization. According to Benner and Tushman (2002, p. 679): “…exploratory innovation involves a shift to a different technological trajectory.”

Baum, Li, and Usher (2000, p. 768) suggest that “…exploitation refers to learning gained via local search, experiential refinement, and selection and reuse of existing routines.”

March (1991) made it explicit that the organization, whose resource and possible investments are limited, can become agile and resilient only if managing well the trade-off between these two innovation capabilities.

Knowledge-based communities (David and Foray, 2003), where individuals interact to innovate collectively in a distributed (von Hippel, 1988) and self-organized mode, are a new organizational form (Romanelli, 1991) that only partially fits the patterns described above. Hence, they provide interesting settings for investigating the connection between individual interactions and community-level capabilities. Community life entails continuous interacting, learning, and achieving (Lave and Wenger, 1991). For classical social scientists such as Goffman (1959) social order and novelty are created at ground level, by individuals, through local practices: “As practices are reproduced over time and across settings, macro-categories [i.e. capabilities] emerge from these interactions and negotiations.” (Powell and Colyvas, 2008, p. 281).

“Community” differs from the traditional notion of “organization” as the former has
emergent means of governance, hierarchies that need to continually be re-legitimized and are often challenged, less formalized, shared goals, and more permeable boundaries (David and Foray, 2003; Mateos-Garcia and Steinmueller, 2008). A community in our interpretation is aligned with Wellman’s (1979) understanding, which highlights that communities are not reducible to a spatial concept (Coleman, 1988). What constitutes a community is interaction among individuals embedded in a social network, shared identity and feeling of belonging (Anderson, 1983), and engagement in a common practice (Wenger, 1998; Brown and Duguid, 1991).

Our focal point is in line with the increasingly popular view in strategic management, organizational learning, and institutional theory that a better understanding of the micro-level of analysis is needed to advance our explanation of macro-categories such as the origins and development of capabilities (Felin and Hesterley, 2007; Argote, 1999; Powell and Colyvas, 2008). In following this avenue of research we are responding to Grant’s (1996) call for an analysis of capabilities at the level of the individual and her activities, rather than an analysis of organizational phenomena. Felin and Foss (2005) proposed a conceptual model, arguing that only through specific studies of the behavior of individuals and their interactions can we gain further insights into how capabilities develop. Consider the following quotes: "To fully explicate organizational anything – whether identity, learning, knowledge or capabilities – one must fundamentally begin with and understand the individuals that compose the whole. ...Truly explaining [...] the organization [...] requires starting with the individual as the central actor." (Felin and Foss, 2005, p. 441). And: “But arguing that individuals are heterogeneous does not imply that the collective level is non-existent or unimportant. Rather, it suggests the importance of explicitly..."
linking the individual and the collective levels." (Felin and Foss, 2005, p. 443). We apply this perspective to our analysis linking the community constructs described above to individual-level behaviors to study the causal relationship at that level. Figure 1 depicts this process in more detail.

The four boxes in Figure 1 show constructs at two levels of analysis: collective and individual. The arrows linking the boxes represent the proposed mechanisms underlying the relationship between constructs (Whetten, 1989). The collective constructs in the model are connected by the dotted Arrow 4, arguing that community capabilities can be explained by the structure of community interaction. Arrows 1, 2 and 3 “explode” this association, first, by mapping the initial set of collective constructs onto the individual level (arrow 1). Arrow 2 then describes the causal relationship in terms of individual behavior linking the focal individual’s actions to the interaction of her peers. Eventually, the result of individual actions is projected back to the collective level as indicated by Arrow 3. We develop arrows 1 and 3 theoretically, and explore the causal link represented by Arrow 2 using data from the research setting described in the next section.

3. Model and research setting

Our empirical setting is an OSS community. OSS is software that is released under licenses that guarantee that the innovation, production, and maintenance of the program are undertaken by a community of individuals, who collaborate in a self-organized way through the Internet, often gathering in projects hosted on dedicated online platforms. This
makes OSS an interesting example of a knowledge-based community, offering insights that may be generalized, mutatis mutandis, to other similar communities.

To date, studies of the OSS phenomenon have mostly focused on the incentives for participation in online debates and the creation of code (Raymond, 1998; Lerner and Tirole, 2002; Ghosh et al., 2002; David et al., 2003; von Hippel and von Krogh, 2003). Despite many attempts (e.g. Lakhani and Wolf, 2005), it is proving difficult to reduce the extreme heterogeneity of motivations into a focused taxonomy. For instance, David and Shapiro (2008) ran a cluster analysis of motivation-related data for more than 1,300 individuals and found that there is no dominant typology of contributors.

These heterogeneous individuals self-select their tasks (Garzarelli and Langlois, 2008) and decide voluntarily how much effort and energy to allocate to them. The governance structure of projects therefore is not given a priori, but changes over time and emerges as the result of individual interactions. Evolution of the social processes of interaction is determined by an open process of discussion and confrontation (Muller, 2006), legitimation (Mateos-Garcia and Steinmueller, 2008), and, sometimes, subsequent formalization (O’Mahony and Ferraro, 2007). Yet these processes do not emerge in a vacuum, as OSS projects are not isolated, one from another. Individuals move between projects and collaborate in different development teams at the same time. This makes projects permeable to influences from other projects, and creates interconnections that can alter the direction and magnitude of their development (Grewal et al., 2006). This leads to the conclusion that the OSS innovation process is undertaken by an intrinsically self-organized social body: the community of interacting OSS developers (Smith and Kollock,
The properties of this complex social body have yet to be investigated in depth. Some work has been done on understanding whether a self-organized community is capable of producing software that can compete with proprietary software in terms of quality or novelty (Klincewicz, 2005; Rossi Lamastra, 2008). Another stream of the literature has investigated the community as alternative or complementary to a system based on the usual intellectual property rights (e.g. Saint-Paul, 2003; Casadesus-Masanell and Ghemawat, 2006; Economides and Katsamakas, 2006). With a few exceptions (Dalle and Jullien, 2003; Gambardella and Hall, 2006), the connection between individual interaction and more aggregated outcomes is confined most to the project level.

As this review reveals, the OSS innovation process couples an irreducible heterogeneity of individual motivations (Belenzon and Shankerman, 2008) with outcomes at community level, for instance capabilities, by means of a self-organizing process centered around individuals’ interactions. The direct consequence of this structure is that, in order to understand how community-level properties - such as the development of capabilities - are shaped, individual actions and social interaction must be at the center of analysis.

We develop this within the conceptual model described in Figure 1 and apply it stepwise to our setting.

Arrow 1. A community is a social body that rests on the consciousness of belonging together, and on the affirmation of a condition of mutual dependence (Tönnies, 1957).
Members of the community ascribe high value to the relationships among them. This means that the social space in which individuals interact is the arena where the community “comes alive”. Distinctions in terms of the media through which interaction in communities occurs have received little attention in the literature. Over time, interaction developed across different types of media evolves and can be classified into genres, which relate to distinct practices, meaning, and knowledge (Yates and Orikowski, 1992). For example, in most organizations we can distinguish interaction between individuals, in terms of being verbal (i.e. coordinating, searching, etc.) or achieved through the undertaking of joint tasks (i.e. performing or contributing to an activity or product).

Scholars applying a community perspective to organizational learning have a clear picture of this process. Amin and Cohendet (2004, p. 80) summarize it in this way: Orr (1996) studied technical representatives at Xerox and he “…found that … [having] breakfast and lunch [together] … allowed the reps to learn through discussion of common problems, narration of experiences … . Similarly, he found that … problems where solved when two reps … came together to work on a machine hitting on a solution after hours of talk.” In self-organized post-bureaucratic organizations, distributed individuals frequently interact using communication technology (Kellogg et al., 2006, Yates et al., 1999). In the literature on OSS, the role of online conversations and “immaterial” artifacts such as software is clearly recognized. Lanzara and Morner (2005) argue that both interaction means are vehicles of the organizational structures and social rules, and crucial coordination mechanisms enabling individuals’ collaboration (Baldwin and Clark, 2006; McCormack et al., 2006; Kuk, 2006; Lee and Cole, 2003).

We decompose the social arena of a community into two main types of interaction
supporting the community structure: interaction through verbal communication and interaction through the exchange of artifacts (i.e. collaboration on tasks) (Desantis and Monge, 1999). In our case, the former is realized through messages exchanged in forums, mailing lists, or chat rooms associated with ongoing projects, through distribution of, and participation in questionnaires, and through other forms of online verbal communication. The latter takes place through the exchange of artifacts that are produced collectively such as bug reports, patch files, feature and support requests, and source code submitted to the projects’ online repositories.

The exchange of artifacts and verbal communication are different in nature. Artifacts relate only to the code produced by the community. Hence, they convey more than the knowledge that is made explicit, but only within the limits allowed by the boundaries of the exchanged object. A patch file, for example, represents the conceptualization of the problem to be solved, while a bug report carries not only information on the problem encountered, but also the cognition of the user who identified it. At the same time, both are related to the specific problem that the patch is meant to solve or the bug report is meant to highlight (Crowston and Howison, 2006), and thus are constrained. Through verbal communication, a concept may instead be formulated in very general terms, overcoming the limits imposed by objects, but becoming more blurred and more heavily dependent on the commonalities between the mental models of the message sender and receiver (Crowston et al., 2005). Interaction developed through artifacts therefore is very different from online exchange of ideas and opinions. Interacting mainly through one or other of the two modes has important implications at the individuals level (e.g. acting on the cognitive structures and mental models of the team members) and at the social level (e.g. with
respect to the capabilities of the team or the community as a collective social body).

The previous discussion allows us to map the first construct at community level to the first construct at the individual level (Arrow 1 in our model): different interaction structures at the level of the community are captured at the individual level considering the propensity to exchange artifacts or to communicate verbally by the collaborators of the focal individuals, that is, by those populating the social space surrounding her.

**Arrow 2.** “Arrow 2’s explanation focuses on the interaction and associated outcomes. For example, March (1991) ... focuses on interaction and socialization as the key determinant to individual and consequently organizational learning” (Felin and Foss, 2005, p. 17).

This quote hints at how the previous individual-level construct developed for Arrow 1 can be related to individual behavior. We can measure the impact of other community members’ propensities to exchange artifacts and communicate verbally on a focal individual’s actions. This is aligned with the argument that the influence of peers in OSS communities plays a crucial role in directing individuals’ activities. Lin (2004) shows that individuals that innovate and exchange OSS artifacts with peers embark on a process of interaction that leads to the construction of a common understanding of the problem to be solved and, through this, to a common identity that shapes their future actions (Wenger, 1998). Elliott and Scacchi (2003) show that online conversations have a clear impact on the way conflicts are resolved and on the decisions made by the individuals involved. Raymond (1998) argued that in OSS, the opinions of peers are crucial drivers of individuals’ actions, an intuition that was confirmed theoretically and empirically by David et al. (2006) and Dalle and David (2005). In other words, individuals developing
OSS are immersed in a social texture that is composed of the actions of their peers. Our claim is that the different means of communication through which this interaction is realized have different impacts on individual behaviors, and on the related community-level outcome: capabilities.

**Arrow 3.** In deciding what community-level capabilities to investigate, we follow March’s (1991) intuition, and limit our discussion to two capabilities tightly linked to innovation: exploration and exploitation. Since the OSS community is considered to be an organization composed of a network of different projects (Grewal et al., 2006), operationally, we claim that when an individual chooses to enter an existing project, her efforts are devoted to an existing enterprise, more resources are allocated to the further development of already existing software, and the community’s exploitation capabilities are improved (David & Rullani, 2008). To define exploration in this setting consider that new OSS projects are rarely based on “new-to-the-world” code (Klincewicz, 2005), and usually are built on existing code (Haefliger et al., 2008). However, their exploratory nature is assured by the fact that they gather a team of individuals that likely is different from the team that developed the original code. For example, when a new project is created as a result of forking⁴, the code is the same for both the “old” and the “new” project. However, the individuals developing the new code left the old project because they had different ideas about how the project should evolve. Hence, the result of forking is a new project that certainly is based on “old” code, but whose intent is precisely to develop something different, along a new trajectory. Therefore, when an individual chooses to create a new project, her energies and time are devoted to the construction of something that points towards an unexplored direction, and community-level exploration
is enhanced (David and Rullani, 2008). Arrow 3 represents this mapping from the individual’s project launching and joining activities and community-level exploration and exploitation, respectively.

With this specification of the arrows, the adapted model comes together as specified in Figure 2.

-------------------------------------------
INSERT FIGURE 2 ABOUT HERE
-------------------------------------------

Hypotheses

Interaction through collaboration in an innovation process, for example, via communication technology, often builds on the content and focus of previous objects (David, 1985). In OSS communities, individuals innovate together, contributing concrete artifacts (i.e. source code). Such interaction goes beyond a global search for original features and novel functionality, and involves suggesting real solutions to real problems. Lee and Cole (2003, p. 644) underline how: “..the Linux self-designing system generates cultural artifacts such as source code….that function as the locus of learning.”. For example, the reuse of knowledge manifested in specific artifacts is a key feature in the process through which individuals in OSS communities develop new source code (Haefliger, 2008). Along the same lines, Lin stresses that these initial conditions of path dependency and reuse were fundamental to the development of Emacs, the program that gave birth to the OSS phenomenon: “Stallman did not sit down and write [Emacs]
immediately [...]. Instead, he looked up the database and found that Mikkelson had made a WYSIWYG feature for TECO. He then integrated his idea into that. If Mikkelson’s work had not existed, we may have seen a different technical option taken, as the ‘problem’ may have been defined differently” (Lin, 2004, p. 7). This is a crucial argument. The condition Lin refers to is related to the existence of a previous program. This very fact was a “constraint” which, while boosting the innovation of Emacs, was implicitly directing its future technological path. This example demonstrates that when individual interaction occurs through the supply or exchange of artifacts, the innovation process aligns closely with the existing product, likely gaining in efficiency through exploitation. We thus propose the following hypothesis:

**Hypothesis 1**: the more that community-level interaction is realized through *artifacts*, the more individuals will engage in community-level exploitation by joining existing projects

On the other hand, and precisely because parts of the solution are already to hand, when interaction takes place through the exchange of artifacts, the propensity to set up a totally different enterprise by founding a new project is less likely. In this case it is not enough to have access to the code and improve it. To attract further individuals, a potential initiator needs to have a vision (Lerner and Tirole, 2002) and a mental model able to interpret the points of view of peers (Crowston et al., 2005) on the potential of a new project. This requires that the initiator be exposed to a debate that is not limited by the continuous problem-solving activity that develops around artifacts. Additionally, such discussion should be able to include more people than the exchange of artifacts, because the details and the characteristics of the topics debated are accessible to many heterogeneous
community members, not just those who posses the skills needed to act on the artifact and understand the “footprints of others” on it. This increase in participants’ heterogeneity offers the potential new-project launcher the possibility to consider her ideas from other – perhaps unexpected - perspectives (Jeppesen and Lakhani, forthcoming). We therefore argue that:

**Hypothesis 2**: the *more* that community-level interaction is realized through *artifacts*, the *less* individuals will engage in community-level exploration by launching new projects.

In OSS communities individuals communicate mainly on line, through forum, chats, questionnaires, and the like. This interaction is composed of messages of text rather than mainly lines of zeros and ones. In other words, the interaction does not have a direct link to the source code and being un-bound from the path-dependent constraints of existing code, interaction through verbal communication has a wider space to explore. Of course, verbal communication is dependent on prior discourse, but is less restricted to following existing paths of thinking. In other words, exchange of code aligns collaborators on trajectories strictly related to the *solutions* of the problem at hand, while verbal communication moves the interaction to the level of the collective *framing* of the problem (Simon 1969). Thus, verbal communication spans broader cognitive boundaries than communication through artifacts. Consistent with this idea, Hargadon and Bechky (2006) found that to capture “the eureka moment” of exploration it is necessary to examine the verbal communication between individuals. Correspondingly, Sutton and Hargadon’s (1999) analysis of idea generation at IDEO shows that verbal communication such as brainstorming is a crucial lever accelerating the creativity process. Thus, verbal
communication creates more room for the conception of new projects that can realize new ideas. Given this, we posit that:

**Hypothesis 3**: the *more* the community-level interaction is realized through online *verbal communication*, the *more* individuals will engage in community-level exploration by launching new projects.

However, precisely because the new idea being implemented goes further than the existing code, joining existing projects is likely to be less frequent. The exchange of artifacts offers problem-specific material, solutions, and conversations. This material is a fundamental source of exploitation, a phase when it is important to be close to the artifact that the individuals are jointly creating. In this phase, too many discussions can easily degenerate into “noise” because they open development paths that are too far removed from the current stage of development. An example of this is Alan Cox’s (1998, online) explanation of why Linux 8086, an existing project that needed further implementation, failed to grow: “The problem that started to arise was the arrival of a lot of ... dangerously half clued people with opinions - not code, opinions. ... How to remove just those who talk and do not do anything is a research topic”. Recalling Linus Torvalds’s motto “Show me the code”, Cox points to the fact that artifacts legitimate participation because they represent a precise contribution to an existing project that needs to grow. A similar point is made by von Krogh et al. (2003) in the case of Freenet. Mapping this onto the community-level of analysis, we propose that:

**Hypothesis 4**: the *more* that community-level interaction is realized through online *verbal*
communication, the less individuals will engage in community-level exploitation by joining existing projects.

Table I summarizes the four hypotheses:

-------------------------------------------
INSERT TABLE I ABOUT HERE
-------------------------------------------

The four hypotheses are neither mutually exclusive nor complementary, as it can be assumed that more resources for exploration, for example, do not decreases the amounts available to be allocated to exploration. Many individuals involved in OSS develop code in their spare time and a large proportion work just few hours a week on their projects (David et al., 2003). Moreover, many projects are small: in January 2003, 67.05% of the projects hosted on SourceForge had only one member (Giuri et al., 2008). Volunteers exhibit a resource slack that increases the flexibility of the innovation process. Debates and exchanges of code may not only direct existing resources to different projects (e.g. Dalle et al., 2009), they may also create new resources attracting new developers and motivating those already on the platform work harder or longer. In our self-organized setting based on volunteers, the trade-off between exploration and exploitation observed in firms may play a role, but does not necessarily do so. Processes such as those considered here may increase (or decrease) the resources for both capabilities at the same time, even disproving or confirming all our hypotheses.
Method and data

To test our hypotheses we gathered data on the activities of individuals working on OSS projects on the SourceForge.net platform (www.sourceforge.net, SF.net henceforth). SF.net provides support for the collective and open development of OSS, and currently is the largest platform in the world: in February 2009 it counted more than 2 million registered individuals and 230,000 registered projects. Individuals registered on SF.net can launch their own projects, browse the platform and find projects that they may be interested in joining, submit bug reports or code patches, suggest new program features, ask for support, participate in discussions on forums or mailing lists, provide feedback and opinions through questionnaires launched by the projects or, more often, by the SF.net staff. All these actions are reported in the dataset, which, in its various versions, has been widely exploited in the literature (e.g. Lerner and Tirole, 2005; Comino et al., 2006)5.

Our version of the dataset allowed us to follow individuals’ activities from September 2000 to December 2002 (i.e. 28 months). We run two Poisson estimations (with fixed effect and robust standard errors) to predict both the number of new projects a focal individual launches in the course of a month, and the number of existing project joined by the same individual in the same period of time. Individuals who never tried to launch a new project or never joined a team are automatically excluded because their influence is fully captured by their fixed effects. In the final version of the dataset each individual is observed from the month of entry into SF.net as a registered user, to December 2002, resulting in an unbalanced panel spanning a maximum of 28 thirty-day periods and composed of a total of 55,916 observations for 2,598 individuals.
The focus of the paper is on how community-level exploration and exploitation are influenced by the features of the community interaction structure. The first set of constructs is captured here by project joining and project launch activities by individuals. We use the variables $JOINED_{it}$, to represent the number of existing projects joined by individual $i$ at time $t$, and $LAUNCHED_{it}$, to represent the number of proposals for new projects filed in the same period by the same individual.

The hypotheses are based on the distinction between two types of interactions: verbal communication and interaction via the exchange of artifacts. We seek to identify the effect of each of these types on the subsequent behavior of the focal individual. This raises the problem of endogeneity: if individual $i$ interacts with the other members of the projects she is involved in, the observed level of interaction (both in terms or artifacts and in terms of verbal communication) will be endogenous. This means we cannot use direct measurement to capture the level of interaction surrounding an individual, and need to rely on indirect measures. Therefore, we created two proxies. For the exchange of artifacts, our data allow us to count the number of bug reports, feature or support requests, patch files and other type of artifacts submitted by participants in individual $i$’s projects to projects in which $i$ does not participate ($INTERACTION\_ART_{it}$). Using the actions undertaken by individual $i$’s collaborators in projects $i$ is not part of, we seek to diminish $i$’s influence on the level of interaction surrounding her, while at the same time retaining the ability to capture $i$’s collaborators’ propensity to communicate through a certain media, and thus also the level and quality of interaction they will adopt towards (and around) the focal individual. Following the same logic, we created the variable $INTERACTION\_VERB_{it}$ counting the number of questionnaires responded to by members of $i$’s projects at time $t$. 

22
When individual \( i \)'s collaborators respond to large number of questionnaires, this is an indication of their higher propensity to interact verbally. Using this variable, endogeneity is reduced by the fact that the focal individual in most of the cases has only a negligible role in creating and promoting questionnaires\(^6\).

To further counter endogeneity, we use fixed effects to account for individual attributes that are unlikely to radically change in the time window under the analysis. Among these, we can include gender, age, nationality, country of residence, education, main skills, sector of employment, family composition and the like. We also introduce a series of time-variant controls. We account first for an individual’s tenure in SF.net counting the number of periods since her registration on the platform \((TENURE_{it})\). We then count the messages sent by the individual to the forums \((FORUM_{it})\) or posted as “news” on a project website \((NEWS_{it})\) in period \( t \) as proxies for the individual’s level of participation on the website. We count the number of projects the individual belongs to in that period \((NPROJ_{it})\) and include a dummy variable to identify whether the individual belongs to the “main project”, that is, that gathering the SF.net maintainers \((MAINPRJ_{it})\). We then move to the aggregate level and count the individual’s collaborators, that is, members of the project teams to which the individual belongs \((NCOLLAB_{it})\). This variable is included in the regressions with \(NPROJ_{it}\), to normalize all the other variables for the size of the social environment in which the individual is immersed. We control for the characteristics of the projects \( i \) participates in. We take into account average registration period \((PROJ\_REGTIME_{it})\), performance based on average number of file releases produced \((FILE\_RELEASES_{it})\), ideological attitude based on the number of projects adopting the most restrictive OSS license, the General Public License \((GPL_{it})\), and their connectivity in the network by
averaging - first inside each project, and then across all the projects - the number of projects participated in by each of i’s collaborators \( (\text{NETWORK}_d) \). Tables II, III and IV report the statistics and the correlations for the variables described in this section.

We also deal with endogeneity by considering that individuals who, over time, become more likely to launch or join a project, might also exhibit a higher probability to attract more communicative or productive developers, stimulate communication or production around them, or move to projects involving more communicative or more productive developers. In fact, we chose the controls described above because they form a combination of variables that captures the endogeneity raised by the three mentioned mechanisms (Rullani, 2008). This use of controls allows for an unbiased interpretation of the impact of the main regressors (Hamilton and Nickerson, 2003, p. 61-62). This technique is enhanced by the fact that since OSS is developed online (see O’Mahony and Ferraro, 2004, 2007 for a discussion) our data depict almost perfectly the world where individuals interact. We can observe all the main factors affecting our regressions (at the least via proxies) and can capture them through controls. We also lagged our regressors one period because each of the three main mechanisms described above is based on a precise sequence of events which can be ruled out by imposing a lagged time structure. Given this, the estimation of the number of projects launched (equation 1) and projects joined (equation 2) by an individual \( i \) in period \( t \) on the basis of certain controls, level of
collaborators, and typology of interaction measured at period $t-1$ is performed using a Poisson model with fixed effects and robust standard errors. This approach is consistent with what Belenzon and Shankerman (2008) proposed in a similar setting. Table V reports the results of the regressions testing significance through two-tailed and one-tailed tests, and transforming the coefficients in incidence rate ratios (IRR). IRR represent the percentage increase in the expected number of projects initiated (for eq. 1) or joined (in eq. 2) with all variables at their means, and are appropriate for assessing the magnitude of the effects of our regressors.

Results

All the results relative to the variables of interest are as expected: when all variables are at their means, the impact of the interaction through artifacts is significantly negative (i.e. IRR<1) on the number of projects launched and is significantly positive (IRR>1) on the number of projects joined. In contrast, $INTERACTION_{\text{VERB}(t-i)}$ has a significant positive impact on projects launched and a significant negative impact on projects joined. All four hypotheses are confirmed.

To further investigate our results it is useful to examine the relationship between the phenomenon (i.e., individual $i$’s exposure to a flow of interaction between collaborators) and the proxies we used to capture that phenomenon ($i$’s collaborators’ artifact exchanges
or questionnaire answers, a *series of distinct events*). We do this by transforming our estimates based on the occurrence of precise events, into estimates that capture the continuous flow of interaction surrounding individual $i$. We can obtain the overall effect of a continuous flow of stimuli over $t$ periods by computing the $t$-th power of each IRR. Using this approach we can demonstrate that after one year of continuous exposure to the exchange of artifacts, a developer launches 12.4% fewer projects, but joins 8.7% more projects. For a six-month period the numbers are 6.4% and 4.2%. If on the other hand the exposure is to verbal interaction, an individual launches 85.8% more projects and joins 33.1% fewer projects (36.3% and 18.2% in 6 months). These numbers are relative to a unit change in the dependant variable when all the variables are set at their means, which for $INTERACTION\_VERB_i$ is 0.239 and for $INTERACTION\_ART_i$ is 1.026 (see Table II). If we consider that the standard deviation of the former is 1.038 and of the latter is 4.212, these effects acquire even more significance.

**Discussion**

Our results provide new insights into how innovation capabilities are accumulated and nurtured in self-organizing knowledge-based communities. In such environments, individuals’ actions, the modes of their interaction, and macro-level constructs such as capabilities, are bound together into a complex system. The structure of interaction within a knowledge-based community, i.e. the media through which participants communicate, and the structure given to these communications, determines how each single participant interacts with her peers. These instances form the texture of the social environment in which each individual is embedded. We show that both the ‘institutional’ thickness (Amin and Thrift, 1995) of this fabric (i.e. the quantity, the amount of interaction), and the
material that constitutes it (the quality, the type of interaction) significantly affect the behavior of the participants in the community. The effects are visible at both levels: individual and community.

At the individual level the results show that the connection between an individual’s innovative behavior and the quantity and quality of the interaction surrounding her, is tighter than is generally assumed in the literature. The possibility to abstract from the path dependencies of existing code and production processes frees the imaginations of participants, allowing them to work together without the constraints imposed by artifacts. This flow of ideas searches for new spaces in the unbounded enterprises of new projects, and creates tensions that expand the technological frontier of the community. When the same interaction takes place through artifacts, the imagination is not free to move far, and participants tend to join existing enterprises and “drill” the technological landscape in the effort to realize a higher level of community exploitation.

A bird’s eye view of these results offers a more in-depth picture. Knowledge-based communities are composite social objects (David and Foray, 2003). In some regions of the social environment of the community interaction is based on verbal communication, for example, in those social spaces where individuals debate concept such as “openness” and “freedom”, so crucial in OSS production and diffusion (Stewart and Gosain, 2006; Dahlander, 2007). Other regions of community interaction are much more closely linked to the code, and the messages individuals exchange are more implicit and embodied in the structure of a bug report or a patch file. Our results imply that different social regions in the community exhibit a different combination of exploration and exploitation capabilities.
Where interaction is based on the exchange of artifacts, the community focuses on the existing projects and developing the existing technology. We cannot expect high levels of exploration from the groups populating these social spaces. However, in those social regions where interaction develops through verbal communication, exploration is the main activity undertaken by participants: new projects are created and expand the existing technological base of the community. The different modes of interaction taking place in the different regions of the community determine which activities the teams of participants embedded in that region will undertake, and thus which technological opportunities the community will explore and which known technological trajectories (Dosi, 1982) will be investigated more in depth.

Our results have also interesting managerial implications. We have shown that the media through which interaction takes place play a fundamental role in framing the actions of individuals and in shaping their self-organized social environment. Framing is crucial in such environments: "Rather than shaping the pattern that constitutes a strategy [...], managers shape the context within which it emerges." (Anderson, 1999, p. 229). Firms engaging with knowledge-based communities can use the insights from this study to design strategic tools to support and frame interactions inside their communities. They should be able to foster the exploration phases through the use of tools that support verbal communication among participants, from face-to-face meetings, to mailing lists, charts, forums, and questionnaires. They can also induce the allocation of more resource to exploitation through the provision of tools and processes that anchor interaction in the artifacts the community is producing. These can be in-person collaboration on the artifacts, as in the case of Orr’s Xerox technicians (Orr, 1996; Brown and Duguid, 1991)
or Wenger’s (1998) claim processing center. They could also be online repositories of artifacts, such as CVS (Lopez-Fernandez et al., 2004) and bug-reporting systems (Crowston and Howison, 2006). Resources may be moved from one capability to another, changing the accessibility, features and general terms of usage of these tools.

A fourth argument emerges from these results. As explained in Section 4, self-organizing knowledge-based communities revolve around the voluntary efforts of participants. Volunteers can join or expand the effort they devote to a certain collective activity if the conditions under which it is undertaken increase their private benefit (von Hippel and von Krogh, 2003). In this sense, the resources the community relies on are flexible, and a particular social process (in our case, interaction) could increase the amount of resources allocated to both exploration and exploitation. This means that in self-organizing contexts, we expect the trade-off between exploration and exploitation observed in firms (He and Wong, 2004) to be less prominent. Our results would seem not to support this idea, but rather to uphold the traditional view that more resources devoted to one area (capability) result in a corresponding reduction in the resources devoted to another. However, there is a different reason for why this is true in our context. In the traditional conceptualization of the balancing act between the two capabilities this situation arises because the overall amount of resources and investments is fixed and cannot be easily increased. In the case of a self-organized community resources are more flexible, but the trade-off between exploration and exploitation remains because the structure of interaction within the community either constrains individuals’ ideas to the limits of the artifact, fostering exploitation and hindering exploration, or allows exploration of new ideas but at the expense of a clearer definition of the problems, reducing exploitation\(^8\). Our results show
that in self-organizing knowledge-based communities the trade-off between exploration and exploitation does not have a resource-based explanation but is *cognitive* in nature (Cohen, et al., 2007).

**Conclusion**

This paper aimed to provide an in-depth, empirically based understanding of how capabilities are developed through individual interaction in self-organizing environments. Not only is this interesting to better conceptualize capability development residing in knowledge-based communities (Brown and Duguid, 1991; David and Foray, 2003; Cole and Lee, 2003), but also it offers important new insights into how the development of capabilities really takes place through individual interactions and actions (Felin and Hesterley, 2007).

Building on the capability literature, drawing from phenomena-based research on the dynamics of OSS communities, and operationalizing the conceptual model in Felin and Foss (2005), we found that the exploration and exploitation capabilities of OSS communities are dependent upon the type and volume of interaction among individuals. In an OSS community individuals populating social groups where the exchange of code and artifacts is the main means of interaction, tend to contribute by joining ongoing projects and thus exploit existing technological trajectories (Dosi, 1982). Those embedded in a region of the community where interaction is mainly verbal tend to explore new territory by initiating new projects. The exchange of artifacts limits the ability jointly to imagine new ventures because it ties discussions to the ongoing development process. Verbal communication hinders exploitation because it moves the discussions too far from the
existing products. The capability trade-off between explorative and exploitative resources
is determined by the cognitive constraints each interaction mode imposes on the
individuals involved.

We arrived at these conclusions via three means. We showed theoretically that the
structure of interaction in the OSS community can be captured at the individual level by
the specific medium (verbal messages or exchange of artifacts) developers use to interact.
We provided arguments to support the idea that joining existing projects and creating new
projects are individual behaviors that lead to community-level exploration and
exploitation. We empirically connected the two individual-level phenomena through
regression analysis on a dataset representing the activities of 2,598 individuals in the
period September 2000 to December 2002 in OSS projects hosted on SourceForge, the
largest OSS development platform worldwide.

Our results should be of interest both to scholars investigating capabilities in new settings
such as self-organizing groups, and researchers trying to understand knowledge-based
communities and the way they organize knowledge creation and sharing. They also have
implications for managers of firms using communities as complementary assets in the
innovation process (Dahlander and Wallin, 2006) and for project leaders in general
(Lerner and Tirole, 2002). Our findings specify how resource allocation in communities
can be directed towards different innovation capabilities by monitoring and adjusting the
tools and media individuals in the community employ for communication and
collaboration on joint tasks.
The paper has some limitations. On the empirical side, better measurement of both types of interaction would be useful. We used the exchange of artifacts as a proxy for individuals’ propensity to use this medium when interacting. Source code provided directly into the code archives of each project is certainly a better proxy, but unfortunately we had no data on this. Similarly, responses to questionnaires and the number of messages individuals post on project forums (which we used as a robustness check) were our only available choices, while mailing lists conversations are probably closer to the phenomenon we want to measure. Nevertheless, we believe we have captured the propensity of individuals to communicate through a certain medium in a way that is meaningful. On the theoretical side, we do not take account of other typologies of interaction, such as face-to-face meetings, or more indirect ways of interacting, such as through message broadcasting. Nevertheless, our insights can be applied to the online communities that exhibit features that by and large are similar to those of the OSS community. We limited our perspective on capability development to the dynamics of resources allocation and accumulation, while more research is needed regarding the mechanisms of capability creation and transformation. Despite these limitations, we believe we have shed new light on the process of capability accumulation in self-organized communities, and its foundations in social interactions and the actions of individuals.
References


Research Policy, 32, 1-11.


International Conference on Software Engineering ICSE.


Organization Science.


(accessed on December 2009).


UK: Cambridge University Press.


Figures

Figure 1. The Relationship Between Collective Constructs and Individual Behavior.

Figure 2. Collective Constructs and Individual Behaviors in the Present Context.
### Table I. Descriptive Statistics (N=55,916).

<table>
<thead>
<tr>
<th>Level of the Analysis</th>
<th>Type of Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual (behavior)</td>
<td>Community (capability)</td>
</tr>
<tr>
<td>Artifacts exchange</td>
<td>Verbal communication</td>
</tr>
<tr>
<td><strong>Project joining</strong></td>
<td>Exploitation</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Project launching</strong></td>
<td>Exploration</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table II. Descriptive Statistics (N=55,916).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOINED$_i$</td>
<td>0.067</td>
<td>0.275</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>LAUNCHED$_i$</td>
<td>0.080</td>
<td>0.307</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>NCOLLAB$_i$</td>
<td>7.909</td>
<td>13.644</td>
<td>3</td>
<td>0</td>
<td>190</td>
</tr>
<tr>
<td>PROJ_REGTIME$_i$</td>
<td>16.501</td>
<td>7.762</td>
<td>18</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>GPL$_i$</td>
<td>1.335</td>
<td>1.194</td>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>NETWORK$_i$</td>
<td>1.866</td>
<td>1.199</td>
<td>1.79</td>
<td>0</td>
<td>17.23</td>
</tr>
<tr>
<td>FILE_RELEASES$_i$</td>
<td>0.291</td>
<td>1.005</td>
<td>0</td>
<td>0</td>
<td>32.5</td>
</tr>
<tr>
<td>FORUM$_i$</td>
<td>0.362</td>
<td>3.132</td>
<td>0</td>
<td>0</td>
<td>205</td>
</tr>
<tr>
<td>NEWS$_i$</td>
<td>0.107</td>
<td>0.585</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>NPROJ$_i$</td>
<td>1.924</td>
<td>1.501</td>
<td>2</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>MAINPROJ$_i$</td>
<td>0.001</td>
<td>0.034</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TENURE$_i$</td>
<td>11.584</td>
<td>6.719</td>
<td>11</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>INTERACTION_VERB$_i$</td>
<td>0.239</td>
<td>1.038</td>
<td>0</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>INTERACTION_ART$_i$</td>
<td>1.026</td>
<td>4.212</td>
<td>0</td>
<td>0</td>
<td>106</td>
</tr>
</tbody>
</table>
Table III. Correlation Matrix (N=55,916).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>JOINEIt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2]</td>
<td>LAUNCHEDit</td>
<td>0.0038</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3]</td>
<td>NCOLLABit</td>
<td>0.0784</td>
<td>-0.0079</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[4]</td>
<td>PROJ_REGTIMEit</td>
<td>0.0651*</td>
<td>0.1254*</td>
<td>0.1403*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[5]</td>
<td>GLPit</td>
<td>0.0862*</td>
<td>0.0791*</td>
<td>0.2309*</td>
<td>0.3815*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[6]</td>
<td>NETWORKit</td>
<td>0.1072*</td>
<td>0.0887*</td>
<td>0.2999*</td>
<td>0.4636*</td>
<td>0.5706*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[7]</td>
<td>FILE_RELEASESit</td>
<td>0.0117*</td>
<td>0.0746*</td>
<td>0.1789*</td>
<td>0.0651*</td>
<td>0.0054</td>
<td>0.0440*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>[8]</td>
<td>FORUMit</td>
<td>0.0039</td>
<td>0.0357*</td>
<td>0.0611*</td>
<td>0.0300*</td>
<td>0.0348*</td>
<td>0.0109*</td>
<td>0.0890*</td>
<td>1</td>
</tr>
<tr>
<td>[9]</td>
<td>NEWSit</td>
<td>0.0076</td>
<td>0.1918*</td>
<td>0.0182*</td>
<td>0.0819*</td>
<td>0.0662*</td>
<td>0.0989*</td>
<td>0.2175*</td>
<td>0.1670*</td>
</tr>
<tr>
<td>[10]</td>
<td>NPROJit</td>
<td>0.1239*</td>
<td>0.1091*</td>
<td>0.3932*</td>
<td>0.4271*</td>
<td>0.6693*</td>
<td>0.8312*</td>
<td>0.0175*</td>
<td>0.0225*</td>
</tr>
<tr>
<td>[11]</td>
<td>MAINPROJit</td>
<td>0.0033</td>
<td>0.0153*</td>
<td>0.0712*</td>
<td>-0.0072</td>
<td>-0.0091*</td>
<td>0.1120*</td>
<td>-0.007</td>
<td>0.0022</td>
</tr>
<tr>
<td>[12]</td>
<td>TENUREit</td>
<td>-0.0103*</td>
<td>-0.1212*</td>
<td>0.1634*</td>
<td>0.3575*</td>
<td>0.2895*</td>
<td>0.3466*</td>
<td>-0.0771*</td>
<td>-0.0257*</td>
</tr>
<tr>
<td>[13]</td>
<td>INTERACTION_VERBit</td>
<td>0.0455*</td>
<td>0.0062</td>
<td>0.4710*</td>
<td>0.0429*</td>
<td>0.0464*</td>
<td>0.2294*</td>
<td>0.1468*</td>
<td>0.0123*</td>
</tr>
<tr>
<td>[14]</td>
<td>INTERACTION_ARTit</td>
<td>0.0375*</td>
<td>0.0859*</td>
<td>0.3864*</td>
<td>0.0736*</td>
<td>0.0810*</td>
<td>0.0754*</td>
<td>0.1756*</td>
<td>0.1388*</td>
</tr>
</tbody>
</table>

* Significant at 5%.

Table IV. Correlation Matrix (N=55,916).

<table>
<thead>
<tr>
<th>#</th>
<th>Variable</th>
<th>[9]</th>
<th>[10]</th>
<th>[11]</th>
<th>[12]</th>
<th>[13]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[9]</td>
<td>NEWSit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[10]</td>
<td>NPROJit</td>
<td>0.1155*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[11]</td>
<td>MAINPROJit</td>
<td>-0.0026</td>
<td>0.0327*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[12]</td>
<td>TENUREit</td>
<td>-0.0366*</td>
<td>0.3650*</td>
<td>-0.0174*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>[13]</td>
<td>INTERACTION_VERBit</td>
<td>0.0051</td>
<td>0.1808*</td>
<td>0.1165*</td>
<td>0.0496*</td>
<td>1</td>
</tr>
<tr>
<td>[14]</td>
<td>INTERACTION_ARTit</td>
<td>0.0948*</td>
<td>0.1026*</td>
<td>0.0498*</td>
<td>-0.0758*</td>
<td>0.2462*</td>
</tr>
</tbody>
</table>

* Significant at 5%.
## Table V. Poisson regression models (fixed effect, robust standard errors).

<table>
<thead>
<tr>
<th>Variable</th>
<th>LAUNCHED&lt;sub&gt;t&lt;/sub&gt; (Incidence Rates Ratio)</th>
<th>JOINED&lt;sub&gt;t&lt;/sub&gt; (Incidence Rates Ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERACTION_ART&lt;sub&gt;(t-1)&lt;/sub&gt; †</td>
<td>0.989 (0.006)**</td>
<td>1.007 (0.004)**</td>
</tr>
<tr>
<td>INTERACTION_VERB&lt;sub&gt;(t-1)&lt;/sub&gt; †</td>
<td>1.053 (0.016)***</td>
<td>0.967 (0.017)**</td>
</tr>
<tr>
<td>NCOLLAB&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>1.020 (0.004)***</td>
<td>1.00604 (0.002)***</td>
</tr>
<tr>
<td>PROJ_REGTIME&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>0.957 (0.008)***</td>
<td>1.05125 (0.005)***</td>
</tr>
<tr>
<td>GPL&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>0.730 (0.083)***</td>
<td>1.166342 (0.072)**</td>
</tr>
<tr>
<td>NETWORK&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>1.637 (0.175)***</td>
<td>1.319138 (0.061)***</td>
</tr>
<tr>
<td>FILE_RELEASES&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>1.041 (0.016)***</td>
<td>1.00258 (0.017)</td>
</tr>
<tr>
<td>FORUM&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>1.002 (0.005)</td>
<td>0.9934104 (0.007)</td>
</tr>
<tr>
<td>NEWS&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>0.992 (0.040)</td>
<td>0.9697362 (0.033)</td>
</tr>
<tr>
<td>NPROJ&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>0.499 (0.066)***</td>
<td>1.005417 (0.062)</td>
</tr>
<tr>
<td>MAINPROJ&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>4.100 (5.672)</td>
<td>2.03201 (1.988)</td>
</tr>
<tr>
<td>TENURE&lt;sub&gt;(t-1)&lt;/sub&gt;</td>
<td>1.015 (0.006)**</td>
<td>0.9280314 (0.004)***</td>
</tr>
</tbody>
</table>

Log Likelihood | -10121.9 | -10469.9 |
Prob > χ² | 0.0000 | 0.0000 |
Obs. (individuals) | 55,916 (2,598) |

† Consistently with our hypotheses, for this regressor we performed one-tail tests. The results remain significant for two-tail tests. For the other regressors we performed directly two-tail tests as we had no prior expectations on the sign of the coefficients.

* = Significant at 10%; ** = Significant at 5%; *** = Significant at 1%.
NOTES

1 We follow Richardson (1972) and take an activity-based view on capabilities: A capability is a capacity of an organization of being good at performing and coordinating a certain activity. We are aligned also with the broad interpretation of the notion offered by Henderson & Cockburn (1994:4) as the abilities of an organization to “deploy its resources and to develop new ones”.

2 Community capabilities have their origins and dynamics located at the organizational level of a community (Wolfe 2007) rather than of an “organization” as identified by Weber (1947) and Blau and Scott (1962). Community capabilities are as Grant (1996:377) puts it: “Integration of specialist knowledge to perform a discrete productive task…”

3 OSS is released using specific licenses that allow any subsequent user of the software to copy, modify, improve and redistribute the original program, but require the application of the same principles to the resulting software. The original code and its subsequent evolutions cannot be made proprietary by anyone. (see Lerner and Tirole, 2005; for a discussion of the different OSS licenses and of how and to what extent each one of them applies these principles). The resulting cooperative regime underlying these licenses spread worldwide. Nowadays, the most diffused software for servers (more than 60% of its market) is Apache, an OSS program.

4 A project experiences a “fork” when part of the team developing it (or some external individual) takes the code and starts to improve it separately from the rest of the team. This results in two distinct projects that are however initially based on the same code.

5 The reader can refer to Howison and Crowston (2004) and Comino et al. (2005) for an overview of the problems connected to the dataset, while we refer the reader to the paper by David and Rullani (2008) to see how the data have been cleaned and treated to assure consistency and reliability.

6 We also used forum messages, filtered considering the number of projects i’s collaborators send messages to. Results are confirmed, even if the coefficient relating INTERACTION_ART_{it} to LAUNCHED_{it} remains negative but turns non-significant. Being a forum an open arena for debates, it is evident that this measure is subject to a higher degree of endogeneity than the number of questionnaires answered by member of i’s projects.

7 The estimated have been computed in STATA using xtpqml, fe, a command developed by Tim...
Simcoe. We run a Haussmann test, that confirmed the preference for fixed effect. Overdispersion was handled using robust standard errors rather than estimating Negative binomial models (Hausman et al., 1984) because if the conditional variance is wrongly specified, Negative Binomial models lead to inconsistent estimated of the conditional mean, while Poisson models lead to consistent estimations of the conditional mean of the distribution even with over-dispersion (Belenzon and Shankerman, 2008). Nevertheless, we estimated also Negative binomial models with fixed effect, and results were confirmed. It was not possible to find a procedure able to estimate the same model with zero-inflation. To have an idea of the possible effect induced by this problem, we reduced the dependant variables to a dummy equal to 1 if the original variable was greater than 0, and 0 otherwise. We then run two logistic regressions with fixed effect and judge if distinguishing only between zeros and non-zeros outcomes may have lead to different results. We obtained results perfectly in line with the main regressions presented in the paper. Finally, the estimation of the two equation separately is justified by the correlation between $JOINED_{it}$ and $LAUNCHED_{it}$, which is non significant (and significant but equal to 0.0098 when one period is lost due to the lag). The same argument is applied by Giuri et al. (forthcoming).

8 We run the same regressions represented in Table V including an interaction term between $INTERACTION_{VERB_{it-1}}$ and $INTERACTION_{ART_{it-1}}$. This term was always negative but never significant.