FIRM-SPONSORED DEVELOPERS IN OPEN SOURCE SOFTWARE PROJECTS: A SOCIAL CAPITAL PERSPECTIVE

Research in Progress

Homscheid, Dirk, Institute for Management and Institute for Web Science and Technologies, Koblenz, University of Koblenz-Landau, Germany, dhomscheid@uni-koblenz.de

Schaarschmidt, Mario, Institute for Management and Institute for Web Science and Technologies, Koblenz, University of Koblenz-Landau, Germany, mario.schaarschmidt@uni-koblenz.de

Staab, Steffen, Institute for Web Science and Technologies, Koblenz, University of Koblenz-Landau, Germany, staab@uni-koblenz.de

Abstract

The way in which companies benefit from open source software (OSS) communities varies and corresponds with the business strategy they maintain. One way of establishing influence in OSS communities is by deploying own resources to an OSS project. Assigning own paid developers to work for an OSS project is a suitable means to influence project work. On the other hand, the pertinent literature on user communities and governance in OSS maintains that a large proportion of influence individuals have in a community depends on their position in the community. This view is reflected by social capital theory, which posits that strong relationships and network positions that are advantageous to access information are valuable resources that affect different downstream variables, most importantly value creation. Thus, this study aims to extend research that has used social capital theory to investigate online communities by testing a conceptual model of social capital and individual’s value creation and assessing the influence of firm-sponsorship on the context.

Keywords: Social Capital, Network Ties, Open Source Software, Linux Kernel Community.
1 Introduction

Open source software (OSS) communities consist of users, developers as well as firms and constitute a resource pool companies may utilise to complement their own resource base (Grand et al., 2004). The way in which companies benefit from OSS communities varies and corresponds with the strategy they maintain (Dahlander and Magnusson, 2008). For example, software development firms can benefit by capturing technical knowledge available in the OSS community (Colombo et al., 2013) but also cut costs through integrating OSS in their own products. These different strategies become possible as OSS shares characteristics of a public good, that is, non-excludability and non-rivalry (Wasko et al., 2009).

Although these characteristics enable firms to appropriate from OSS as such, it is difficult to differentiate own offerings that rely on, or benefit from OSS from those of competitors (Da Silva and Alwi, 2008). Thus, it is vitally important to at least influence the development trajectory of an OSS project to optimise the benefits that arise from making use of OSS communities (Schaarschmidt, 2012). One way of establishing influence in OSS communities is by deploying own resources to an OSS project (Schaarschmidt et al., 2013; West and O'Mahony, 2008). Firm-sponsored developers are contractually bonded to the firm and are likely to behave in the firm’s intention while contributing to the OSS project as “a man on the inside” (Dahlander and Wallin, 2006). Thus, assigning own paid developers to work for an OSS project is a suitable means to influence project work.

On the other hand, the pertinent literature on user communities and governance in OSS maintains that a large proportion of influence individuals have in a community depends on their position in the community (e.g., Crowston and Howison, 2006; Dahlander and O'Mahony, 2011). This view is reflected by social capital theory, which posits that strong relationships and network positions that are advantageous to access information are valuable resources that affect different downstream variables, most importantly value creation (Tsai and Ghoshal, 1998). Recently, the relation between network position in a community and different positive outcomes has been emphasised in various areas. For example, Chou and He (2011) were able to show that a developer’s social capital positively affects expertise integration, that is, individuals with high social capital synthesise project-related information for other community members. Relatedly, Wasko and Faraj (2005) found for electronic networks of practice – communities that share characteristics with OSS communities – that social capital is associated with knowledge contributions. Aside from these research efforts, important aspects that pertain to the role of sponsorship have not been addressed yet. Research has to show whether the associations between network position and positive outcome as predicted by social capital theory are independent of developers’ profession or are different for heterogeneous developer groups. Against this background, this study aims to extend research that has used social capital theory to investigate online communities by addressing the following central research question:

*How is the relation between an OSS contributor’s social capital and his/her created value affected by firm-sponsorship?*

To approach the research question, we adopt and test a conceptual model of social capital and individual’s value creation. Our study has theoretical as well as managerial implications. Theoretically, this study sheds much-needed light on the effect of sponsorship in online communities on social capital on an individual level. Managerially, this study reveals mechanisms that help firms engaged in OSS development to influence value creation. This research-in-progress paper is organised as follows. First,

---

1 In this study the terms *developer* and *contributor* are used synonymously to denote people who are active in OSS projects and in Linux kernel development.
we give an introduction to the social capital concept and define OSS communities. Second, we provide our research model along with a detailed derivation of the hypotheses. We then describe our research approach and close by discussing the expected contribution of this research.

2 Theoretical Background

2.1 The Concept of Social Capital

Over the last decades, a broad range of researchers investigated the concept of social capital. A variety of contributions arose from these investigations involving disciplines such as political sciences (e.g., Putnam et al., 1993, Putnam, 1995a, 1995b), sociology (e.g., Burt, 1992, 1997; Coleman, 1988, 1990; Portes, 1998), economics (e.g., Glaeser et al., 2002; Knack and Keefer, 1997; Woolcock and Narayan, 2000) and management (e.g., Adler and Kwon, 2002; Inkpen and Tsang, 2005; Kwon and Adler, 2014; Nahapiet and Ghoshal, 1998). However, the debate on the issue of social capital is not entirely new. The insight that social capital has a positive influence on the functioning of communities dates back to Lyda J. Hanifan (1916), who referred to social capital as the first author in the academic literature.

In consequence of the mentioned contributions from a variety of research fields, many definitions for social capital arose over the last years, as the social capital concept was applied to and modified for different research contexts (Adler and Kwon, 2002). This richness can be a pitfall, as the various definitions also include different ways to conceptualise and measure social capital which in turn leads to diversity in evaluating the causal mechanisms in the macro- (collective) and micro- (individual) processes (Lin, 1999).

In this paper we draw on the social capital view of Lin (1999; 2001), as he focuses in his considerations of social capital theory on the causal relationship between the position of an individual in a network and the associated outcomes due to the access to resources. In this sense, Lin defines social capital as “the resources embedded in social networks accessed and used by actors for actions” (Lin, 2001, p. 25). Embedded resources in social networks may lead to positive outcomes because of the following four reasons. First, an individual will benefit from the flow of information if s/he has social ties with people in strategically or hierarchically relevant positions. The associated opportunities in consequence of the information flow would not be available without the corresponding relations (Chou and He, 2011). Second, if an individual is centrally positioned in the social network, s/he will be able to influence decisions by exploiting social ties. Third, other people will attribute more and higher social credentials to an individual, if s/he can show off relations to strategically positioned actors. Lastly, social relationships are seen as reinforcements of one’s identity and recognition, which provide emotional support as well as public acknowledgement (Lin, 1999).

Research pertaining to social capital distinguishes at least two levels of analysis; an individual and a group level. On the individual level, the focus is on access and usage of resources embedded in social structures to obtain returns in purposeful actions, such as getting a more favourable job. On the group level, the focus is on development and maintaining social capital as a collective asset (Lin, 1999). Here, we draw on individual-level aspects of social capital.

2.2 Open Source Software Community

In general, a community arises when different people come together and share a common interest (Preece, 2000). Virtual communities (e.g., OSS communities) are social associations, which develop by means of and through the Internet, when people find virtually together to share opinions about topics of common interest and therefore build social relations over time (Rheingold, 2000). OSS development communities share community characteristics described by Whittaker et al. (1997) as there are
shared goals such as providing superior software to users (O’Mahony and Bechky, 2008), repeated participation, accessibility of shared resources (Wasko et al., 2009), generalised reciprocity, and shared language (Stewart and Gosain, 2006). Consequently, Von Hippel and Von Krogh (2003) conceptualise OSS development communities as “Internet-based communities of software developers who voluntarily collaborate in order to develop software that they or their organizations need” (Von Hippel and Von Krogh, 2003, p. 209). Besides the fact that OSS communities consist of hobbyists, who voluntarily provide their resources to the community, the definition also involves another important contributor group – organisations. Organisations differ from hobbyists in terms of their motivation to engage and are represented in the community by their employed developers. In turn, employed developers might be considered as proxies for firm interests in the community. However, until now, the majority of research efforts pertained to communities characterised by homophilous actors. With firm-affiliated individuals, the community consists of heterogeneous actors who have not yet been in a research focus to date.

2.3 Hypotheses Development

Nahapiet and Ghoshal (1998) introduced a conceptual model, in which social capital – divided into structural, cognitive and relational dimensions, hereafter also referred to as structural, relational and cognitive capital – is the prerequisite to build up and harness intellectual capital in organisations. This conceptual model was taken by Tsai and Ghoshal (1998) to derive and empirically examine a model of social capital and firm’s value creation. We draw on these elaborations to conceptualise and evaluate a model of social capital and individual’s value creation. In this study, the focus and thus the unit of analysis, respectively, is on the individual level, involving individual contributors and their relationships to other contributors in OSS communities. Our research model is depicted in Figure 1 and will be explained through the development of our hypotheses in the three following subchapters.

Figure 1. Research Model

2.3.1 Relations of social capital dimensions to each other

The structural dimension of social capital reflects the structure of relations between individuals in a group or community (Granovetter, 1992). The structural attributes of social capital are defined on the basis of the network structure, which is spanned by the individuals, including network ties, network configuration and appropriable organisation (Nahapiet and Ghoshal, 1998). These attributes, belonging to the context of an individual’s social interaction, are seen to shape an individual’s trust in relationships to others, to influence the norms, the obligations and expectations or the identification of an individual. The latter named characteristics are connected to the relational dimension of social capital, which comprises assets that are anchored in relationships (e.g., trust) (Nahapiet and Ghoshal, 1998). Thus, individuals form trusting relationships over time with their community peers through social in-
interactions. This circumstance was confirmed by previous studies, for example Granovetter (1985) and Tsai and Ghoshal (1998). Following the aforementioned coherence and applying it to the context of OSS contributors, we propose that the formation of trustful relations also takes place through online social interaction, which leads to the following hypothesis:

**H1:** Structural capital is positively associated with a contributor's relational capital.

The cognitive dimension of social capital stands for a shared language and for shared codes as well as shared narratives among individuals in a group or community (Nahapiet and Ghoshal, 1998). These cognitive elements are essential because they support a common understanding of joint objectives in a social system and they thereby simultaneously facilitate the formation of trusting relationships among individuals active in that social system (Barber, 1983; Nahapiet and Ghoshal, 1998). Due to collective goals and values individuals in a community do more likely trust one another, as they assume that they are all pursuing the same objectives (Tsai and Ghoshal, 1998). The same holds true for virtual or online communities, where individuals not physically, but digitally, through the Internet, find together and share a common interest and common goals (O’Mahony and Bechky, 2008), as well as use a common language (Stewart and Gosain, 2006). Based on these contexts, we establish the following hypothesis for OSS contributors:

**H2:** Cognitive capital is positively associated with a contributor's relational capital.

The relation between the structural and the cognitive dimension of social capital is based on the assumption that social interaction is vital for individuals to form and to pass on common objectives and values in a group or community (Tsai and Ghoshal, 1998). With focus on contributors adopting objectives, languages and values through their interaction in an OSS community, we conclude the following hypothesis:

**H3:** Structural capital is positively associated with a contributor’s cognitive capital.

### 2.3.2 Relations of social capital dimensions to contribution

As described above, the structural dimension of social capital is based on an individual’s social ties and its social interactions. Tsai and Ghoshal (1998) describe social ties as pathways through which information and resources flow. In particular, through social interaction an individual may obtain information and resources from other individuals through unofficial channels. Previous studies have shown that this circumstance is all the more true, the more central an individual is in a community (e.g., Crowston and Howison, 2006; Dahlander and O’Mahony, 2011). The fast and easy access to information and resources enables an individual to perform better in accordance with the objectives of the group or community. Studies on this subject researched on the influence of an individual’s group centrality, which is associated with the structural dimension of social capital, on knowledge contribution (e.g., Wasko and Faraj, 2005), expertise integration (e.g., Chou and He, 2011) or research impact (e.g., Li et al., 2013). In the sense of the aforementioned studies we define source code contribution in our research model as proxy for value created by individuals in a community. As a high centrality of an individual in a community can lead to a greater volume and speed of resource flows (Li et al., 2013), we propose that central contributors in a OSS community can gain advantages from structural capital and obtain more information, knowledge and resources. The consequence of the locational as well as information advantage in the community is likely to be reflected in a higher contribution to the source code. These considerations lead to the following hypothesis:

**H4:** Structural capital is positively associated with a contributor's source code contribution.

Trust, norms, obligations and identification are key concepts of the relational dimension of social capital. Individuals can achieve more valuable relational capital by maintaining strong relationships with other individuals in a group (Hooff and Huysman, 2009; Nahapiet and Ghoshal, 1998). Accordingly, strong and trustworthy relationships are seen as a major asset for cooperation, resource acquisition and
knowledge sharing in virtual communities (Chang and Chuang, 2011; Ridings et al., 2002). Moreover, a high level of trust among individuals in a community can foster joint problem solving and commitment to as well as identification with the community (Wasko and Faraj, 2005). If an individual maintains strong relationships to his peers and he is fully committed to the community work, that person is most likely willing to spend more time on fulfilling tasks for the community (Hsu and Hung, 2013). Drawing on these findings, we establish the following hypothesis for the context of committed contributors in OSS communities:

H5: Relational capital is positively associated with a contributor's source code contribution.

The cognitive dimension of social capital consists of a shared language and a shared vision, as previously described. Both characteristics form and strengthen over time and enable individuals to engage in a meaningful exchange of ideas, interests and knowledge within a collective (Nahapiet and Ghoshal, 1998). In the same vein individuals expand their knowledge and expertise during the process of exchange with their community peers (Wasko and Faraj, 2005). Such procedures enhance the cognitive capital of individuals and will increase the likelihood that involved individuals will contribute more, for example knowledge (Chiu et al., 2006; Wasko and Faraj, 2005), to the community. We transfer the existing findings to the context of OSS contributors and their contribution to the source code and conclude the following:

H6: Cognitive capital is positively associated with a contributor's source code contribution.

2.3.3 Sponsorship as moderator

OSS communities receive contributions not only from hobbyist, but also from firms, more precisely from employed developers contributing to the community on behalf of their employer (Grand et al., 2004; Schaarschmidt and Von Kortzfleisch, 2009). As a result, today’s successful OSS projects obtain contributions from hobbyists, universities, research centres, as well as from software vendors and user firms (Homscheid et al., 2015; Teigland et al., 2014; Schaarschmidt and Von Kortzfleisch, 2015). As firms pursue certain objectives with their dedication of employed developers to an OSS community, it is comprehensible that they may influence the development trajectory of an OSS project to get desired returns from their engagement (Schaarschmidt, 2012). Consequently, firm-sponsored contributors will get involved in a different way in the community than hobbyists (Dahlander and Wallin, 2006). To assess if there is any difference in the relation between social capital and source code contribution for different contributor groups, especially hobbyist and firms-sponsored developers, we propose the following three hypotheses. These hypotheses are backed by the idea that being sponsored provides access to a wider set of resources (Dahlander and Wallin, 2006) that complement resources provided through social capital.

H7: Sponsorship positively moderates the relation between a contributor’s structural capital and his/her source code contribution.

H8: Sponsorship positively moderates the relation between a contributor’s relational capital and his/her source code contribution.

H9: Sponsorship positively moderates the relation between a contributor’s cognitive capital and his/her source code contribution.

3 Research Approach and Method

3.1 Research Context

To find a relevant OSS project for our research, we have taken different aspects into account (e.g., size of the project, activity and continuity, company involvement, availability of a large set of data). Final-
ly, we chose the Linux kernel (LK) project as our research context, which has served as an example for OSS in many previous studies (e.g., Lee and Cole, 2003). The LK project was initiated by Linus Torvalds in 1991 and has been one of the most active OSS projects since its beginning. There are software releases every three months on average, which are possible because of the fast-moving development process and the broad foundation of contributors, ranging from hobbyists to companies. Thus, it involves more people than any other OSS project. The kernel itself makes up the core component of any Linux system and is used in operating systems for mobile devices right up to operating systems for supercomputers. Typically, a new release of the kernel comprises over 10,000 patches contributed by over 1,100 developers representing over 225 companies and is published under the GNU General Public Licence v2 (Corbet et al., 2013). Besides the fact that the LK is one of the largest cooperative software projects ever started, it has also an economic relevance, as many companies have business models that rely on the LK or on software working on top of the LK, respectively. Many of these companies do actively participate in the improvement of the kernel and thereby take effect on the orientation of the development. Very active companies in the kernel development, among others, are RedHat, Intel, IBM, Samsung, Google and Oracle (Corbet et al., 2013).

3.2 Measures

The measures for this research can be segmented by their source of data. We will analyse network data with means of social network analysis for the period from 2005 to 2014 as well as LK source code data from 2005 to 2014. The two sources of data limit the possibility that our results are affected by common method variance (Podsakoff et al., 2003). In the following paragraph, the operationalisation of the research model is described.

Network Data (2005 – 2014)
The needed network data was captured from a web archive, which displays the LK mailing list. To measure social capital the three social capital dimensions developed by Nahapet and Ghoshal (1998) are utilised as follows.

Structural Capital. In this study we follow prior research (e.g., Li et al., 2013) and use social network analysis (SNA) methods to measure structural capital by different centrality aspects of an individual, that is, degree centrality, closeness centrality, betweenness centrality and eigenvector centrality (Borgatti et al., 1998) in the LK community.

Relational Capital. For the measurement of relational capital we follow the approach of Wasko and Faraj (2005) and use SNA for examining tie strength and the distribution of messages over time of a LK contributor, as a measure of commitment. Further, reciprocity is measured through the analysis of dyadic relations of the developers following the approach of Squartini et al. (2013).

Cognitive Capital. In previous studies cognitive capital was mostly measured through a questionnaire (e.g., Chang and Chuang, 2011; Chou and He, 2011; Tsai and Ghoshal, 1998). As we want to determine cognitive capital from the network data, we measure cognitive capital by using topic models (Steyvers and Griffiths, 2007). In a first step, we detect different topic groups within the LK mailing list by the aid of topic modelling. In a second step, we analyse how well the individual developers fit with their emails into the existing topic groups. As a shared language is indispensable for the formation of cognitive capital and as the exchange of ideas, knowledge and expertise in a community support this (Chiu et al., 2006; Wasko and Faraj, 2005), higher probabilities of a developer to fit with

2 http://marc.info/?l=linux-kernel
3 http://vger.kernel.org/vger-lists.html#linux-kernel
his/her conversation into more than one topic group within the LK community are related with higher cognitive capital.

Sponsorship. The identified individuals interacting in the LK mailing list are heterophilious as some act on behalf of companies while others might be considered hobbyists. To get a deeper understanding of whether the individuals in the mailing list are affiliated with a firm or not, the domain name of the email addresses was used to assign people to a contributor category. Developers sending messages from a domain indicating that the person is employed by the corresponding company are classified as employed contributors, whereas people using email addresses from public email providers such as yahoo.com were classified as hobbyists. Likewise, developers with email addresses indicating universities and research institutions were identified. Assigning LK developers to a contributor category was done in a semi-automated and semi-manual process in order to obtain a high accuracy of the attributions. Detailed information about the different contributor categories is provided in Table 1.

<table>
<thead>
<tr>
<th>Contributor Category</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Companies</td>
<td>People with email addresses from companies</td>
<td>intel.com, redhat.com</td>
</tr>
<tr>
<td>Hobbyists</td>
<td>People with private email addresses</td>
<td>gmail.com, yahoo.com</td>
</tr>
<tr>
<td>Universities</td>
<td>People from universities</td>
<td>columbia.edu, duke.edu</td>
</tr>
<tr>
<td>Research Institutions</td>
<td>People from research institutions and public authorities (e.g., IEEE, government, military)</td>
<td>ieee.org, nasa.gov, army.mil</td>
</tr>
</tbody>
</table>

Table 1: Coded Contributor Categories.


Source Code Contribution. To measure the contribution variable, we calculate a delta in source code contribution on a yearly basis for each individual active in code contribution. All needed data for this calculation was obtained from the LK Git repository.

4 First Results and Expected Contribution

As a first result we provide descriptive information about the LK mailing list dataset in the period from 2005 to 2014. The cleaned dataset comprises 1.94 million messages sent to the main “linux-kernel” mailing list. These messages result from over 567,900 message threads with more than 44,500 messages writers involved. Data continually increased from 10,600 messages (in average) per month in 2005, up to 23,100 messages (in average) per month in 2014.

This research contributes to knowledge in the field of social capital theory in the context of relationships in OSS communities. Furthermore, to the best of our knowledge the operationalisation of all three social capital dimensions through network measures has not been previously considered by researchers. This new form of operationalisation will lead to new knowledge about measuring social capital with the aid of network measures.

Besides the theoretical implications, the intended results of this study are also relevant for management practice. Thus, the results will help to advance IT managers’ understanding of how to interact with a development community in case the resulting software is a public good. Specifically, this research will give insights how to position employees in an OSS project to influence the projects trajectory without having direct control over or authority in the project.

The next research steps involve combining the LK mailing list data and LK Git repository data in one database. Then, using ordinary least square regressions and zero inflated negative binominal regressions, we analyse statistically how different social capital dimensions affect code contribution over time. Finally, we map our results with the pertinent literature on OSS development and social capital.
References


