



# Stipendium „Essays on Communities“ by Inna Smirnova

Endbericht | Call 13 | Stipendium ID 3844

Lizenz: CC-BY

# Inhalt

1	Einleitung.....	3
2	Allgemeines.....	4
3	Ergebnisse .....	6
4	Geplante weiterführende Aktivitäten.....	9
5	Anregungen für Weiterführung durch Dritte.....	9

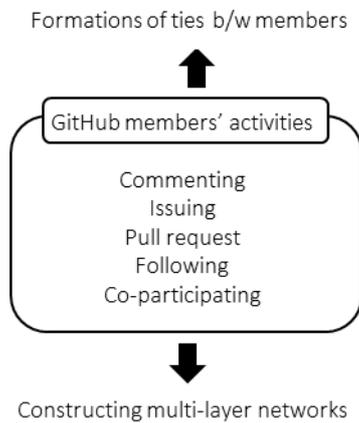
# 1 Einleitung

Innovation is an important source of competitive advantage (Katila and Ahuja 2002, Klevorick *et al.* 1995, Laursen and Salter 2006, Schumpeter 1934). Over the past two decades, a continuous trend toward shifting activity from in-house R&D to the exploitation of externally generated solutions has been observable (Boudreau and Lakhani 2013, Chesbrough 2003, Dahlander and Wallin 2006, Laursen and Salter 2006). In this process, online peer production communities, notably open source communities, whose members collectively work on (technical) problems of joint interest, have become an increasingly important provider of innovative solutions (Dahlander and O'Mahony 2011, O'Mahony and Ferraro 2007, von Hippel and von Krogh 2003). However, as communities are not tied to corporations via standard labor contracts, firms seeking to steer collective activity towards commercially relevant projects need to find alternative ways of incentivizing potential contributors. One opportunity is for firms to (co-)design communities of interest themselves and provide incentive mechanisms that can satisfy the diverse set of motivations and facilitate community participants to invest time and effort in their contributions (Boudreau and Lakhani 2009, Dahlander and Magnusson 2008, Malhotra and Majchrzak 2014, Parker *et al.* 2017, West and O'Mahony 2008). To do so *effectively*, however, firms still need to get an in-depth understanding of how such communities work. This is the broad question I primarily focus on in my PhD dissertation.

As is known today, communities differ from traditional organizations pursuing the goal of producing similar goods/software artifacts in many ways (Dahlander and Frederiksen 2012, Raymond 1999). Notably, contributors do not submit themselves to traditional managerial forms of authoritative task allocation but self-select into projects instead (Puranam *et al.* 2014). In turn, they (mostly) do not receive financial remuneration (O'Mahony and Ferraro 2007). To infer from the absence of direct financial rewards that contributors are purely intrinsically incentivized would be misleading, however. Research conducted to this day conveys a multifaceted picture of what motivates contributors to join public innovation communities on average (Lakhani and Wolf 2005, von Krogh *et al.* 2012). Here, the consensus appears to be that nonpecuniary rewards incentivize outsiders to join and work for the community whereas monetary incentives, in turn, can lead to undesired crowding-out of intrinsic motivation. Interestingly, fulfilling use-needs, gaining recognition and visibility for career-related and unrelated purposes (Jeppesen and Frederiksen 2006, Lakhani and Wolf 2005), personal enjoyment, and satisfying a quest for career-unrelated

learning (Benbya and Belbaly 2010, Lee and Cole 2003) would appear to stimulate individuals working in communities.

Online communities rely exclusively on the informal network structures as major information flow



channels and hence represent an unique multi-layered organizational structure. In my PhD dissertation I examine (data from GitHub platform - largest and most active OSS community) how the division of labour is performed in online communities and how contributors are able to achieve various advantageous structural positions which allow them to increase productivity and enhance their (and their projects') community status. Understanding of how communities actually work and

how they can be efficient helps to advance the current community scholarship and provide a practical advice for organizations on how to direct community production in an efficient manner.

## 2 Allgemeines

*Knappe Zusammenfassung der Arbeit, Fragestellung, Ziel der Arbeit*

### 2.1 PhD Thesis Part I

The first chapter of my dissertation aims at exploring how the task allocation problem is addressed in online peer production communities. We explore which incentives lead skilled as opposed to unskilled contributors to sort into an OSS community to produce output. More specifically, we study how founders of open source projects can not only attract many contributions to their repositories, but instead solicit them from highly skilled individuals who produce high-quality programming code and who may also recommend themselves to firms for future hire as boundary-spanners to help integrate openly produced innovative solutions. This knowledge is critical when seeking to use the community to *both* produce powerful software and act as a screening device for talent. Building on the results from personnel economics that have previously shown a positive relationship between performance-related financial motivation and the attraction of skilled labor (Cadsby *et al.* 2007, Dohmen and Falk 2011, Lazear 2000), we suggest that extrinsic incentives should attract higher-skilled workers more the better such incentives allow contributors to leverage their abilities when attaining their personal goals and minimizing related opportunity costs.

Drawing on the large custom-tailored dataset merging information from two major software community archives—Stack Overflow and GitHub, we show that project owners can identify best coders by growing their projects, being generous in accepting incoming code contributions, and providing (fast) feedback. Our findings indicate that the output of an online community is not (entirely) exogenous to the community owner and that certain project design choices might lead to skill-based sorting patterns across different groups of contributors joining the online community. By studying skill-based sorting of coders on community projects, our research thus also provides a key contribution to the debate on how sharing code may aid in attracting skilled coders and potentially approaching them for hire (Leppämäki and Mustonen 2004).

This project is successfully finished and is under the 1<sup>st</sup> round revision at the MIS Quarterly journal.

## **2.2 PhD Thesis Part II**

In OSS communities, participants can both launch their own projects and manage them as repository founders and make voluntary (helping) contributions to the repositories of others by writing code patches, making improvement suggestions, and proposing novel ideas. Interestingly, these activities, while not mutually exclusive, often tend to be filled out by different individuals (due to time and resource constraints) and lead to some role specialization. Here, it would be typical for repository founders as owners of their OSS projects to manage incoming contributions and ongoing discussions among project contributors, and to make final code acceptance or rejection decisions – rather than to still write much code themselves. Why exactly such a structure would emerge and successfully persist is not trivial, though. Researchers have drawn on a diverse set of theoretical lenses in the past, however, two perspectives have featured most prominently in the explanations: (1) OSS as networks of helping behaviors and (2) OSS as goal-directed systems. According to the first and overall dominating understanding, online communities arise from the exchange between volunteer contributors who adhere to norms of reciprocity (Faraj and Johnson 2011, Levine and Prietula 2013, Simpson *et al.* 2018). Preferential attachment and community project founders' centrality in a network of contributors (Yang *et al.* 2013, Andriani and McKelvey 2009, Grewal *et al.* 2006, Singh *et al.* 2011) have demonstrably engendered exchange patterns between contributors that are compatible with key tenets of social network theory (Burt 1992, Uzzi 1997). According to the second view, these collectives display features of goal-directed systems that actually resemble traditional (for-profit) organizations which build on the division of labor and integration of effort perspectives (Arazy *et al.* 2014, Lee *et al.* 2017, Puranam *et al.* 2014). Yet, each one of the views on its own is struggling to predict how OSS collectives grow and how efficient

growth patterns actually arise. In this study, we seek to reconcile both perspectives by arguing that OSS communities create efficient division of labor that promotes further project growth and success through different types of reciprocal behaviors performed by a project founder towards her contributors.

## 3 Ergebnisse

### *Beschreibung der erreichten Ergebnisse*

In this section I summarize the theoretical predictions we have made for the second dissertational study and our current findings.

#### **3.1 PhD Thesis Part II: theoretical predictions**

We develop founder-level predictions to test our idea of reciprocity as a driver of OSS repository growth. From the founders' standpoint, their code repositories grow when receiving more contributions from other participants. In our study, we propose that in open innovation communities to induce contributions and grow their own projects, project founders have two approaches resting upon the reciprocity principle:

- One approach assumes that contributors reciprocate their work to the focal founder for her contributing to other founders' projects. This lies on the principle of reciprocal behaviour - a founder A sends code to the user B and gets the same favour back.
- The other approach assumes that contributors reciprocate the focal founder for her efforts on community assembly to acknowledge, coordinate and integrate incoming contributions. This is the case of a non-equivalent reciprocity when a founder A receives coding help from a user B but pays back by doing a different activity/favour instead of sending a code patch to B's project.

Both approaches require founders to exert different types of effort in a community. Due to limited resources and time, project founders need to formulate resource allocation strategies and pick one of the available approaches in order to ensure their projects' growth. We suggest that the second approach is more beneficial for the following three reasons.

First, the second approach promotes division of labor and co-specialization and therefore enhances efficiency and productivity. This allows founders to focus on maintenance and contributors on bringing their novel ideas to a project.

Second, the principle of reciprocity embarked in the second approach operates more effectively. When deciding which project to join, contributors need to find out whether a particular founder

is a free rider or not. One of the ways is to actually observe what founder does in a community and on her projects specifically. It seems hard for complete outsiders to assess the complete community-wide activity record of a given founder, however it is fairly easy to observe founder's behavior on a given project. If founder decides to focus on the second approach, her potential contributors are more likely to be certain that their efforts would not be wasted and all their works will be integrated into the chosen for contribution repository (this could be called reciprocity in the community-specific context).

Third, contributors in open innovation communities are interested in enhancing skills and acquiring knowledge as one of the reasons of why they join OSS and certain projects. If founders take the first approach, the feedback to and appreciation of contributors' work would be delayed or poorly provided. The second approach, on the other hand, would allow founders to focus on her contributors' needs first. This will more likely motivate existing contributors to stay on a project and put more effort (studies show that absence of feedback significantly reduces employee/contributor motivation to continue working on a given task/job in general).

### **3.2 PhD Thesis Part II: empirical results**

We tested our predictions on the contribution data from GitHub – an online OSS community that is the largest and most active of its kind. The data contains all activities (project launching, sending and receiving code patches, opening and resolving issues, commenting on different issues raised on a project) performed by registered developers on the platform from the GitHub's inception date (29 October 2007) until June 2011 (the size of the archival data is large and therefore we focused only on a representative sample by restricting the years of our observations). The data gave us more than 60,000 observations of activities (aggregated on a founder-month base) across more than 90,000 GitHub projects.

Our identification strategy includes the deployment of a refined degree centrality measure for code-exchange ties (=pull requests) stipulated between OSS repository founders and other programmers. Here, we link the diverse reciprocal behaviors that a repository founder provides to her community participants and other GitHub founders with her projects' success. For our tests we deployed OLS log-log models with founder level fixed effects to control for time-invariant founder characteristics, which we do not observe (such as race, mother tongue, formal education). We also control for various founder and contributor level factors such as platform tenure, programming performance, overall founding activity, and issue and commenting activities.

Our baseline tests using GitHub individual contribution and project-level data provide us with support for our hypotheses. In line with our predictions, we find that founder's focusing on maintaining her own project and satisfying needs of her contributors helps it to grow further. On the contrary, we find that an effect of equivalent reciprocity is negative (opposite direction from the predictions from the social networks literature), suggesting that GitHub is more of a professional, goal-directed community with norms of division of labor prevailing (rather than a social network like YouTube or Twitter). These findings show that communities (i.e. OSS projects in particular) can be efficient and thus can provide complementary assets to firms' traditional innovation activities.

The next question that then arises is - which communities can be more efficient and bring the greatest value to firms, what are the boundary conditions which define such projects, or sub-communities? Our further theoretical investigation was to test whether different regimes in norm enforcement across projects within the GitHub community could explain differences in projects' efficiency levels (and ultimately amount of output being produced). The expectation is that the projects with higher norm enforcement would have the baseline effects exacerbated. Our tests reveal that the baseline effects we obtained would be stronger the higher the number of community members who closely monitor a project or difference in tenures among project members (i.e. proxy for internal cohesion within a project team) is.

Overall, we find a robust evidence for our predictions. However, as every empirical study is imperfect, we needed to perform a few robustness checks in order to be able to claim the validity and generalizability of our results. Here are a few points/limitations which needed to be addressed:

- 1) There might be an issue with the construct validity for our variable that captures founder's managing activities on her own projects. Generally speaking, project management and maintenance work consists of two crucial parts - managing incoming contributions and providing feedback to contributors. Our measure, however, only includes the former since we had no data records available on the latter commenting part. Theoretically, this omission could distort our results. However, practically, based on this data observation, we can only assume that founders do not consider this activity as being particularly important for the project. In fact, in our related interview studies, a lot of project founders mentioned that this activity takes away their valuable time and they'd rather delegate it, if this was possible.

2) For our final data sample we had to exclude inactive GitHub accounts and empty repositories. For robustness, we reran all analyses with such zero records - results are largely in line with our main findings.

3) Our sample might still be considered a little old. Perhaps, this might in part explain why we do not observe a lot of pull request activity. We tested our hypotheses on a subsample of newer data - results remain unchanged.

This further investigation provided us with more confidence regarding the validity and robustness of our findings.

## 4 Geplante weiterführende Aktivitäten

### *Kurze Angabe von noch offenen Aktivitäten (Zeitplan)*

The main part of the second dissertational chapter is completed. The theory section is structured, and the promising literature direction and research gap are identified. Moreover, GitHub data was successfully collected, and empirical analyses (incl. robustness checks) are performed. Also, tables with main results and illustrations with effect sizes are produced. To sum up, based on the original time-plan of the study, all phases till Phase 3.1 are completed.

As a next step, we are currently combining all materials together to prepare the complete manuscript draft. We are trying to craft the contributions section to position our study favourably in the body of existing literature streams. Our findings contribute to the broader literature on communities (we add the discussion on efficiency of communities to the existing knowledge), organization theory (we advance the current understanding of division of labour in organizations with no authoritative task allocation) and social network theory (we show that different types of reciprocity affect growth/success of a project differently). As a further step, we need to prepare the manuscript and send it out for feedback to receive improvement suggestions. This phase (Phase 3.2 in the original time-plan) is planned to be performed in October-December 2019. After the draft is finalized, we expect to submit the paper for consideration for publication in one of the main Management journals (expected submission date is December 2019).

## 5 Anregungen für Weiterführung durch Dritte

### *Welche Weiterführungen für Dritte ergeben sich durch Ihre Arbeit?*

My dissertation (and the second study in particular) provides several contributions to theory and practice. Focusing on open source communities, my dissertation is going to empirically

demonstrate: 1) how an intra-organizational structure of a GitHub community emerges and develops over time as the projects grow and attract more developers, and 2) how the type of an emerged network is associated with the individual developer's and overall project's output and status within the GitHub community. The latter finding is going to indicate that the output of an online community, which has increasingly been viewed as an important resource that for-profit firms can draw on for their R&D operations, is not (entirely) exogenous to the organizational designer (=project founder in this context). I believe that our models of how people voluntarily form teams and work together towards producing innovations will help organizational scholars and practitioners to learn how to leverage certain network formations in order to raise productivity and community growth and extract value from open source community networks. Moreover, to social network scholars, we bring an idea of non-equivalent reciprocity and suggest that networks evolve differently when participants reciprocally exchange different kinds of resources. Social network research often makes an implicit premise that ties between network participants represent exchanges of same kinds of resources and thus only provides one-sided predictions. Our theory and findings add complementary view/knowledge to this research stream that creates opportunities for further investigation.

## References

- Andriani, P. and McKelvey, B. 2009. Perspective—From Gaussian to Paretian thinking: Causes and implications of power laws in organizations. *Organization Science*, 20(6): 1053-1071.
- Arazy, O., Nov, O., and Ortega, F. 2014. The [Wikipedia] World is Not Flat: on the organizational structure of online production communities.
- Benbya, H. and Belbaly, N. 2010. Understanding developers' motives in open source projects: a multi-theoretical framework. *Communications of the AIS*, 27: 589-610.
- Boudreau, K. and Lakhani, K. 2009. How to manage outside innovation. *MIT Sloan Management Review*, 50(4): 69–76.
- Boudreau, K. and Lakhani, K. 2013. Using the crowd as an innovation partner. *Harvard Business Review*, 91(4): 60–69, 140.
- Burt, R. S. 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.
- Cadsby, C. B., Song, F., and Tapon, F. 2007. Sorting and incentive effects of pay for performance: An experimental investigation. *Academy of Management Journal*, 50(2): 387-405.
- Chesbrough, H. 2003. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business School Press, Boston, MA.
- Dahlander, L. and Frederiksen, L. 2012. The core and cosmopolitans: A relational view of innovation in user communities. *Organization Science*, 23(4): 988-1007.
- Dahlander, L. and Magnusson, M. 2008. How do firms make use of open source communities? *Long Range Planning*, 41(6): 629–649.

- Dahlander, L. and O'Mahony, S. 2011. Progressing to the center: Coordinating project work. *Organization Science*, 22(4): 961-979.
- Dahlander, L. and Wallin M.W. 2006. A man on the inside: Unlocking communities as complementary assets. *Research Policy*, 35(8): 1243–1259.
- Dohmen, T. and Falk, A. 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review*, 101(2): 556-590.
- Faraj, S. and Johnson, S.L. 2011. Network exchange patterns in online communities. *Organization Science*, 22(6): 1464-1480.
- Grewal, R., Lilien, G.L., and Mallapragada, G. 2006. Location, location, location: How network embeddedness affects project success in open source systems. *Management Science*, 52(7): 1043-1056.
- Jeppesen, L.B. and Frederiksen, L. 2006. Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science*, 17(1): 45-63.
- Katila, R. and Ahuja, G. 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal*, 45(6): 1183-1194.
- Klevorick, A.K., Levin, R.C., Nelson, R.R., and Winter, S.G. 1995. On the sources and significance of interindustry differences in technological opportunities. *Research Policy*, 24(2): 185-205.
- Lakhani, K. and Wolf, R. 2005. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In Feller, J.R. et al. (Eds.), *Perspectives on Free and Open Source Software*: 3–21. Cambridge, MA: MIT Press.
- Laurson, K. and Salter, A. 2006. Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. *Strategic Management Journal*, 27(2): 131-150.
- Lazear, E.P. 2000. Performance Pay and Productivity. *The American Economic Review*, 90(0): 1346-1361.
- Lee, G.K. and Cole, R.E. 2003. From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization Science*, 14(6): 633–649.
- Lee, S., Baek, H., and Jahng, J. 2017. Governance strategies for open collaboration: Focusing on resource allocation in open source software development organizations. *International Journal of Information Management*, 37(5): 431-437.
- Leppämäki, M. and Mustonen, M. 2004. Signaling and screening with open source programming. Working paper, Helsinki School of Economics, Helsinki.
- Mallapragada, G., Grewal, R. and Lilien, G., 2012. User-generated open source products: Founder's social capital and time to product release. *Marketing Science*, 31(3): 474-492.
- Malhotra, A. and Majchrzak, A. 2014. Managing crowds in innovation challenges. *California Management Review*, 56(4): 103–123.
- O'Mahony, S. and Ferraro, F. 2007. The emergence of governance in an open source community. *Academy of Management Journal*, 50(5): 1079-1106.
- Parker, G., Alstyne, M.V., and Jiang, X. 2017. Platform ecosystems: How developers invert the firm. *MIS Quarterly*, 41(1): 255–266.
- Puranam, P., Alexy, O., and Reitzig, M. 2014. What's “new” about new forms of organizing? *Academy of Management Review*, 39(2): 162-180.
- Raymond, E. 1999. *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Sebastopolous, CA: O'Reilly & Associates.
- Schumpeter, J.A. 1934. *The Theory of Economic Development*. Harvard University Press, Cambridge, MA.
- Simpson, B., Harrell, A., Melamed, D., Heiserman, N., and Negraia, D.V. 2018. The Roots of Reciprocity: Gratitude and Reputation in Generalized Exchange Systems. *American Sociological Review*, 83(1): 88-110.

- Singh, P.V., Tan, Y., and Mookerjee, V. 2011. Network effects: The influence of structural capital on open source project success. *MIS Quarterly*: 813-829.
- von Hippel, E. and von Krogh, G. 2003. Open source software and the “private-collective” innovation model: Issues for organization science. *Organization Science*, 14(2): 209-223.
- von Krogh, G., Haefliger, S., Spaeth, S., and Wallin, M.W. 2012. Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quarterly*, 36(2): 649–676.
- West, J. and O’Mahony, S. 2008. The role of participation architecture in growing sponsored open source communities. *Indust. Innovation*, 15(2): 145–168.
- Yang, X., Hu, D., and Robert, D.M. 2013. How microblogging networks affect project success of open source software development. *Proc. 46th Hawaii International Conference on System Sciences (HICSS)*, 3178-3186.