The More the Merrier? The Effect of Size of Core Team Subgroups on Success of Open Source Projects

Leonard Przybilla [1], Maximilian Rahn[1], Manuel Wiesche[1], and Helmut Krcmar[1]

[1] Technical University of Munich, Chair for Information Systems, Garching, Germany
{leonard.przybilla,wiesche,krcmar}@in.tum.de
ga89rux@mytum.de

Abstract. Open source software (OSS) has become an important organizational form of building software. Given the desire to understand drivers of OSS project success and the known importance of social structure for team functioning, we investigate the effects of the relative size of contribution-based subgroups on community size of OSS projects. Drawing on extant research on OSS and faultline-based subgrouping, we investigate the relation with project community size of the relative size of subgroups based on reputation, issue focus, contribution extent and contribution persistence. While in several instances non-significant, results suggest a differential relation in which a large share of core members with high reputation, issue focus and persistent contributions positively relate to community size, whereas a large share of extensively contributing members in the core team is negatively related. Our findings are of value to research and practice by furthering the understanding of work in OSS projects.

Keywords: Open Source Software, Subgroups, Community Size, Team Governance.

1 Introduction

Open source and related concepts such as libre or free software development (in the following summarized as Open Source Software or OSS) have gained much traction in the beginning of the century [1] and continue to garner research attention recently [2].

Since most of the members of OSS projects contribute during their spare time and without monetary remuneration, the questions what motivates people to join, to contribute over longer periods of time and how such informal communities are managed have emerged as topics of research. Such issues are all the more relevant since despite overall success of OSS, a large majority of projects is defunct and not maintained [3]–leading to the issue how success in projects can be propelled.

OSS development is characterized as a virtual, distributed form of teamwork in which theoretically anyone can contribute [4]. This implies that developers are likely to differ on a number of attributes. OSS team members' motivations have been found

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

to be manifold–ranging from personal gain such as programming knowledge [1] or reputation [4] to philanthropic intentions [1]. Moreover, OSS team members embrace a common set of specific values and attitudes, which directly relate to work practices [5]. As may be expected from the clique-like description of OSS team interactions, the onboarding process of new members can be riddled with challenges [6].

The notion of positive effects of diversity in members on problem-solving [7] is only to some extent replicated in OSS [8], leading to the questions when and how diversity is conducive to outcomes. Based on characteristics shared by some members of a team and thus separating them from others, diversity can lead to so-called faultlines [7], which in turn may lead to perceivable subgroups [9]. Contingent on the specific reason for formation and configuration of faultlines and subgroups, the direction, i.e. enhancing or harmful, and strength of effects may differ [10, 11].

OSS teams as inherently open entities with a diverse set of members harbor much potential for faultlines and subgroups. Despite the critical importance of joint work in OSS, subgrouping and the resulting configuration as influential phenomena in general group research have, to the best of our knowledge, not received any research attention. We take a first step to addressing this void by investigating the following research question: Does the configuration of contribution-based subgroups in OSS teams relate to success as indicated by community size?

We first provide background information on extant research on OSS as well as faultlines and subgrouping before discussing specific implications in the context of OSS. We then introduce our hypotheses and the method used before reporting results of analysis and discussing implications. Lastly, we provide concluding remarks.

2 Background

In the following, we will briefly introduce extant research on OSS development and subgrouping before discussing the implications of subgroups in the context of OSS.

2.1 Work in Open Source Software Development

The success of OSS is astounding given its organizational challenges. Howison and Crowston (2014) report that organizing OSS is especially difficult for at least three reasons: Challenges presented by distributed work are exacerbated by relying on volunteers, which renders traditional incentive mechanisms ineffective. In addition, the work undertaken in OSS is complex with the associated difficulties [12]. Against the backdrop of previous research on the personality of developers [13], these assertions give rise to the questions which mechanisms help achieve valuable outcomes and why developers join and continue participating in OSS projects in the first place.

The motivations to join OSS projects are manifold. Given its characterization by voluntary contributions [1] and the ensuing absence of monetary remuneration, other causes such as personal motivations prevail. As private benefits, personal need for the developed functions, fun derived from working on the task and learning are key [1].

Membership in the community, the ability to gain reputation, and the possibility to receive job offers are also recurring themes [4]. The strong sense of community is mirrored in OSS participants sharing a common set of beliefs and values [5].

Assuming members assemble in a project, the issue of how work is organized arises. While OSS can be compared to several paradigms of work organization, unique differences are highlighted. Due to its inherently distributed nature, OSS can arguably be related to such teams, albeit results on governance may not be directly transferrable [12]. The voluntary and thus indeterminate nature is mirrored in elements that OSS development shares with agile projects [14]. More testament to the specific type of work accomplished in OSS is given by structural investigations. Typically, a relatively small core of developers contributes the majority of work, which is augmented by the smaller contributions of peripheral members [15]. Considering team composition, strong network ties of developers have been observed to bolster success [16]. For embeddedness of developers and projects, differential effects on success have been observed [17]. Moreover, the proficiency of projects at either developing new features or improving upon existing code has been observed to depend on the structure of collaboration [2].

The presence of a strong sense of community coupled with findings that a core of developers contributes differently than a periphery of developers gives rise to the question on how the configuration of the core relates to success in the larger community. To determine such possible effects, we propose to draw on faultline and subgroup theory.

2.2 Faultlines and Subgrouping

Diversity, i.e. differences in team members regarding attributes such as gender or functional background is found to be conducive to performance in teams by enabling the integration of diverse viewpoints [7]. In OSS, diversity of members has been found to improve some but not all outcomes [8].

Effects of diversity can be explained by so-called faultlines: Latent divisions among members based on characteristics shared by only some [7]. If perceived by members, faultlines are activated and lead to subgroups [18], i.e. several smaller entities within the overarching work teams [11]. For the purpose of this research, the term “subgroup” refers to activated faultlines and is rooted in faultline theory– notwithstanding its use in other contexts.

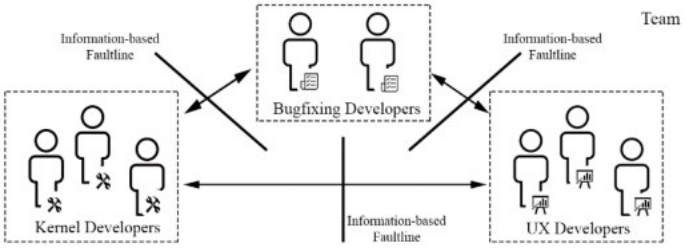

Figure 1: Example of Subgroups based on Information Processing, adapted from [19]

Faultlines and to a larger extent active subgroups have been found to affect team outcomes [9]. Recently, it has been proposed that the reason for subgroup formation may lead to different types of subgroups with different internal processes and thus different effects on group outcomes [11]. Identity-based subgroups due to e.g. differences in age are expected to trigger mostly negative processes, resource-based subgroups due to e.g. status differences harbor the potential for conflict but can boost efficiency, and information-based subgroups due to e.g. different expertise can engender team effectiveness by supporting information processing across groups. Given the need to coordinate knowledge in software development [20], the implications of information-based subgroups could be especially positive. Figure 1 exemplifies the emergence of subgroups based on information-triggered faultlines. The number of subgroups and their balance in terms of membership size, i.e. equally split versus imbalanced subgroups, also influence subgroup effects with e.g. an imbalanced configuration of geographically dispersed members leading to negative effects [10, 11]. Empirically, a complex interaction of subgroup formation, configuration and team outcomes has been observed [21]. In particular, software engineering practices may change subgrouping and its effects [19, 22].

2.3 Faultlines and Subgroups in Open Source Software

Considering the importance of commonly held values and community in OSS [5] and the observed effects of subgrouping raise the question whether harmful or positive effects of subgrouping occur in OSS. To this end, we provide an initial, non-exhaustive assessment of faultline types in OSS development.

By communicating through electronic means, members of OSS projects have limited possibilities to observe characteristics of their peers [8]. Faultlines based on *demographic attributes* may not be perceivable and thus irrelevant–unless members include demographic information in their public profiles. In fact, demographic attributes have not been found to be prominent among members [23].

Motivations to join OSS projects are manifold and thus harbor potential for splitting groups along identity-based faultlines. It could, however, be the case that like-minded individuals cluster in homogeneous groups. Motivations have been found to differ also based on project characteristics, i.e. size [24], which then would attract a specific type of developer. Since membership may not be fully determined by a single motivating factor and projects may cater to more than one need, e.g. enabling learning and at the same time providing opportunities to build reputation, motivation is likely to lead to identity diversity and thus faultlines in OSS projects.

Experience in OSS development in general and the specific project is expected to present an information-based faultline. Differences in professional experience are documented as faultlines [25], additionally in the context of OSS distinct differences in knowledge, which arguably is related to experience, are described [8].

Reputation as an individual's social status is important in the social fabric in OSS [26]. Reputation as the congruence of promised actions and actual behavior [27] is multi-faceted such that positive views in technical aspects can be coupled with negative social evaluations. Given this multidimensionality and basis for authority

[8], reputation is likely to differ for an individual between projects and for individuals within one project. These differences are expected to lead to a hierarchical structure and thus resource-based subgroup [11].

Differences in *activity type* are well-established differentiators in OSS projects and thus a likely faultline item. First, users and developers differ, where users mostly consume and at most make small contributions such as bug reports or minor changes, whereas developers contribute all major code advancements [8]. Within the set of developers, a hierarchy consisting of a core and more peripheral developers has been described: A set of core developers has a disproportionate share of contributions, which entails more influence and reputation, whereas a large number of peripheral developers contributes relatively little code [15].

The *extent and persistence of contributions* is another potential faultline. The overall amount of activity is expected to be an influential member characteristic. Abstracting from the specific contribution behavior, the core-periphery structure of OSS projects [15] is based on the extent of contributions. Activity is an antecedent to previously discussed characteristics such as experience and hierarchy. In addition, by contributing continually, members can build knowledge, which is a key criterion for advancing to more central roles [15]. Drawing on research into other open collaborative processes, roles are expected to be identifiable but flexible over time [28]. A subsequent reduction in activity may thus demote members from the core to peripheral contributors.

3 Hypotheses

We propose a set of hypotheses to investigate the correlation of the relative size of subgroups based on high reputation, issue focus, high contribution extent, and high contribution persistence and success of OSS projects as defined by community size.

3.1 Success in OSS

Success in OSS is not dependent on a single characteristic. The multi-faceted nature of OSS success is evident from the proposition of frameworks to assess success based on diverse indicators [29]. Following previous application [29], we use the size of the non-core OSS project community as an indicator of its external success since contributions of peripheral members are valuable to maintain the project [8]. The onboarding mechanism, i.e. to integrate new developers into the project has been described to be a difficult issue in OSS [6]: Community size is thus apt to indicate how well a project cannot only garner attention but recruit contributors, who potentially can advance to the core team. Based on the preceding discussion of characteristics prone to lead to faultlines and subgroups, we derive hypotheses on the relations of a selected subset of bases for subgrouping that are deemed relevant for community size.

3.2 Hypotheses on the Configuration of Subgroups in OSS Core

For the hypotheses on the relations of the relative size of subgroups defined by faultlines, we draw on findings concerning the stable yet dynamically changing roles in open collaboration [28]. Discussions on the type of potential subgroups and their related effects draw on the typology suggested in [11].

Reputation has been described as an individual's social status in OSS projects [26]. Drawing on extant research in subgrouping, differences in reputation can be related to hierarchical differentiation, which can lead to negative outcomes as resource-based subgroups [11]. At the same time, without perceptions of unfair distribution, a hierarchy can facilitate information processing and thus aid group performance [30]–which relates to potentially positive information-based subgroups. In the context of OSS, reputation has been found to increase trust and satisfaction in members [27, 31]. Considering virtual teams, trust in turn has been observed to increase participation and community activity [32]. Reputation may also facilitate the progression from observing user to contributor with decision power [8] through satisfaction, which leads to participation intentions [27]. In addition, since we focus on the size of the peripheral community as dependent variable, the presence of high-reputation individuals in the project core may signal credibility [27], which may help to attract new members.

H1: The relative size of the subgroup of high-reputation individuals in an OSS project will be positively correlated with community size.

Issue Focus: Different activity backgrounds lead to the potential of information-based subgroups, which can be positive [11]. A large share of issue-focused contributors, i.e. with most activity in creating and commenting on issues, is expected to foster community size. Reporting issues is a known pathway to transition from user to core contributor [33] since issue reports require less specific technical and project-related knowledge than code contributions. With commenting also being part of issue focus, a large share of issue-focused members implies many members may still be starting out as contributors or many people are helping others into the project by sharing knowledge through comments. Core members commenting on issues of newcomers is comparable to mentoring, which has been found to aid onboarding [34]. In addition, a large subgroup based on such behavior may send positive signals of a collaborative culture to outsiders and consequently make the project more attractive.

H2: The relative size of the subgroup of issue-focused individuals in an OSS project will be positively correlated with community size.

Contribution Extent: While rather general, the extent of contributions in projects is expected to foster success and to generate follow-up activity. Similar to reputation, past contributions in a project act as an outside signal of activity and maintenance–as opposed to a majority of OSS projects that are effectively abandoned [3]. Such signals may sway outsiders to become part of the peripheral network. Activity in and of itself is positive in OSS, which is witnessed by an emphasis on practical work in core beliefs [5]. The importance of activity for community building is mirrored in the finding that updates on activity are a key reason for following other members [35]. In addition, for acquiring new casual contributors, a large share of highly active

contributors can make it easier for newcomers to identify who to turn to and ask questions and whose work to study to overcome issues related to a lack of replies found in onboarding [6]. In this sense, a large share of members in a high-activity subgroup may foster efficient processing as a knowledge-based subgroup [11].

H3: The relative size of the subgroup of extensively contributing individuals in an OSS project will be positively correlated with community size.

Contribution Persistence: Analogously to contribution extent, we expect a large share of persistent developers to aid community size. The presence of persistent contributors shows a project is actively developed and thus increases its attractiveness. Past activity may inform future activity [36] and thus benefit future contributions. In addition, persistent developers may be easier to identify and the likelihood of responses are increased, which can address the onboarding issue of receiving no reply from core members [6].

H4: The relative size of the subgroup of persistently contributing individuals in an OSS project will be positively correlated with community size.

Control Variables: To control for systematic differences in OSS projects, we include project age and the existence of previous releases as control variables. Project age is used to control for lifecycle aspects [37] and to capture related effects such as integration in the OSS community, access to resources and progress [38, 39]. As a binary control, the previous existence of releases is used to control for projects that while being actively developed do not declare official releases.

4 Method

4.1 Sampling

Data on OSS projects was obtained from GitHub Archive and a copy of the GHTorrent data on Google BigQuery. We included projects that had at least 100 pull requests or at least 500 commits and at least 2000 comments between January 1st 2014 and August 31st 2017, yielding 6037 projects of which we drew a 10% random sample. Controlling for name changes, the sample contained 580 projects. Since success is expected to be the result of collaboration we followed extant OSS research [37] and applied a lagged structure: Independent variables are collected from a six month timeframe in the middle of the project lifetime, community size in the following six months and control variables in the preceding six months. Sample size was reduced to 482 based on a sufficient level of activity in the reference period and the removal of two outliers showing an extreme level of activity not representative of the majority of projects.

4.2 Operationalization of Measures

Community size as a measure of external interest and thus success, see section 3.1, is operationalized as the extended development community [29] and implemented as the count measure of individuals associated with the project without being part of the

project core. *Reputation of developers* is calculated as the prestige actor proximity index [40] by measuring how connected and how close an individual is to other members. This is a more elaborate approach than the one used by [26] to assess OSS reputation. Links between members are inferred by analyzing the sequence of users' comments, their quotes and direct references in discussions. The index increases with the number of reachable developers and if developers, which are directly or indirectly connected, get closer. *Contribution extent* is operationalized as an individual's share of overall project activity during the period of investigation, in terms of comments, issues, commits, and pull requests. This measure is inspired by previous constructions of developer-activity pairs in networks [2]. *Persistence* of an individual is operationalized as the share of periods with activity in all periods since the individual's first contribution. Following previous research classifying contribution types in OSS [2], *issue focus* shows the relative share of issue-related activity in an individual's contributions. It is operationalized as the ratio of the number of issue-related contributions to a project, e.g. issue reporting and commenting, compared to an individual's overall contributions to the project. We were, however, unable to reliably distinguish comments from issues to those to pull request, which may skew results.

Table 1 Operationalization of Measures

Measure	*Operationalization*	*Calculation*
Community size [Success]	Extended Development Community of p as sum of non-core members i_p^{nc}	$S_p = \sum i_p^{nc}$
Reputation	Connectedness and closeness of developer i to other developers j through comments, quotes, and direct references	$P_{ip} = \frac{I_i/(g_P - 1)}{\Sigma_j d(j,i)/I_i}$ I_i Number of nodes reachable from i g_P nodes in project p $d(j,i)$ distance of j to i
Contribution Extent	Share of i of overall activity in project p	$CE_{ip} = \frac{C_{ip}}{\Sigma_i C_{ip}}$ C_{ip} *Number of Contributions of i to p*
Contribution Persistence	Share of periods with activity since initial activity.	$CP_{ip} = \frac{A_{ip}}{S_{ip}}$ A_{ip} *# of periods with activity of I in p* S_{ip} *# of periods since first contribution of i to p*
Issue Focus	Share of issue-related activities in an individual's contributions	$R_{ip}^{issue\ focus} = \frac{C_{ip}^{issue}}{C_{ip}^{total}}$ C_{ip}^{issue} *# of issue-related contributions of i to p*

4.3 Analysis

For the independent faultline measures, we restricted analysis to the core project based on an activity threshold of twenty contributions. Since the count-based raw scores are prone to project-specific skews, values of faultline measures are normalized first. Drawing on previous findings concerning the specific effects of the relative size of subgroups [10], we operationalize the size of theoretically derived subgroups as the share of team members deviating more than half a standard deviation from the project median. Core members of projects can thus belong to either the high or low-value subgroup on the respective measure. This calculation is done for each independent variable. With the dependent variable being a count measure, Negative Binomial Regression (NBR) has been chosen as regression method.

5 Results

Table 2 Correlations of Variables

	High Reput.	**Issue Focus**	**Contrib. Extent**	**Contrib. Persistence**	**Proj. Age**	**Comm. Size**
High Reputation	1					
Issue Focus	-0.21***	1				
Contrib. Extent	-0.29***	0.29***	1			
Contrib. Persistence	-0.13***	0.16***	0.20***	1		
Project Age	-0.03	0.07	0.10**	0.26***	1	
Community Size	0.09*	0.00	-0.17***	0.10**	0.08*	1
Note:	*p<0.1; **p<0.05; ***p<0.01					

Correlations of variables–shown in Table 2–are relatively small with the maximum value being .29 in absolute terms. The direction of correlations is, however, worth mentioning: the share of members with high reputation is negatively correlated with all other measures, whereas all other correlations are positive. Table 3 details the regression models. The first model only includes the subgroup measures as independent variables, whereas the second and third one add the control variables project age and whether there have been releases in the project. The control variables do not change the direction of correlations but influence levels of significance. While we expected the attribution of members to subgroups to be meaningful for community size, the information content of the model is rather low judging from the pseudo R^2 values.

H1 regarding the effect of a large share of members with high reputation is partly supported: We observe a positive albeit insignificant and small relation.

H2 regarding the effects of a large share of members with a focus on issue activity is likewise partly supported with a small positive, albeit insignificant relation.

H3 regarding the effect of a large share of extensively contributing members is not supported with a highly significant and strong negative relation.

H4 regarding the effect of a large share of persistently contributing members is supported with a significant and strong positive relation.

Table 3 Results of Regression Models

	Community Size (Number of Non-Core Members)		
Model	**1**	**2**	**3**
Relative Size of Subgroup of Members with			
High Reputation	.621*	.411	.413
Issue Focus	.377	.307	.317
Contribution Extent	**-2.176*****	**-1.926*****	**-1.959*****
Contrib. Persistence	**1.283*****	**1.304*****	**1.201*****
Releases		**.491*****	**.484*****
Project Age			.001
Constant	**4.797*****	**4.482*****	**4.171*****
Pearson Dispersion	1.257	1.228	1.26
Pseudo R^2 (McFadden)	0.007	0.011	0.011
Pseudo R^2 (Nagelkerke)	0.076	0.117	0.121
Observations	482	482	482
Log Likelihood	-2,762.340	-2,751.393	-2,750.319
theta	.804*** (.046)	.834*** (.048)	.837*** (.048)
Akaike Inf. Criterion	5,534.680	5,514.786	5,514.639
Note:	*p<0.1;	**p<0.05;	***p<0.01

6 Discussion

Based on the characterization of work in OSS projects and theory on faultlines and subgrouping, we investigated the relation of contribution-based subgroups and community size as a measure of success. While the applied regressions explain only a small share of variance, large sets of core members with high reputation, a focus on issues, and especially persistent contributions positively relate to community size. A large share of extensively contributing members is significantly negatively related.

As expected, a large share of high-reputation core members has a positive but small and after including controls insignificant relation with community size. With reputation being a key aspect of OSS culture [5], we expected the resulting

differences in power, resource access and status to be attenuated by the culture of OSS work and thus to lead to a positive relation. Results suggest, however, that reputation may also in OSS projects lead to negative repercussions–possibly due to an identity- or resource-based subgroup. This relation may interact with the hypothesized positive effect. Operationalized as a social proximity measure, the positive finding is in line with previous work on the positive effect of internal cohesion for OSS success [37]. Structurally speaking, a larger share of developers with more direct access to other core members positively relate to community size. Drawing on previous research describing a positive effect of loosely coupled, decentralized developers for design as opposed to technical work [2], the expected signaling effect of high-reputation projects may draw in peripheral contributors–whereas the technical work of closely related core contributors may trigger feelings of inaptitude and thus present barriers to onboarding [6, 41]. Previous findings thus help interpret the only partly expected findings. In addition, reputation based on the distance of the social network may not be perceivable to outsiders and thus reduce the expected signaling effect. The current operationalization of reputation may overestimate values of members being in constant exchange with others without adding value to the project. Investigating other metrics, e.g. formal collaborator status, thus seems worthwhile.

The positive relation between a larger share of members focusing on issue activity and community size is relatively small and insignificant. With a grain of salt, this result may be interpreted as slight proof of the proposition that issue-focused core members foster community size as defacto mentors helping to overcome onboarding issues [34]. Moreover, supporting others as a core value in OSS [5] could propel membership. This line of reasoning has, however, to be questioned since the correlation between the share of high-reputation and issue-focused individuals is slightly negative. The small effect size may be due to our specific threshold values for considering members part of the project core as it could have included too many contributors and thus left no room for outside community. In addition, as stated before, the operationalization of issue focus suffers from the inability to classify some comments. Effects may be more reliably tested if the content of contributions was to be analyzed in more detail: In particular, the community building effect may be identifiable if responses to activity by non-core members were studied in particular.

The relations of contribution extent and persistence are somewhat surprising: Persistence is–as expected–significantly positively related, while extent is significantly negatively related. The correlation between the two measures is weak, which implies they capture distinct contribution types. Persistence might not only be the sum of contributions over time but may signal activity, future maintenance, and thus value in contributions to outsiders. It might also imply technical proficiency and learning forming part of the OSS culture [5]. Seeing sense in one's contributions can be related to the intrinsic motivational factors as key drivers of OSS membership [1, 4]. Persistence might thus signal a project is worthy of contributions. Extensive contributions on the other hand are operationalized by the overall activity of individuals. A large share of extensively contributing core members might create the impression of a closed circle and thus deter contributions–relating to the finding that newcomers face barriers in where to start contributing [6]. In addition, the negative

correlation of contribution measures and high reputation is noteworthy since it may imply that contribution quality by high-reputation individuals is distinct from quantity.

This research provides initial evidence that the relative size of subgroups in OSS projects may have differential effects based on underlying faultlines. Findings on positive relations add to existing research on the onboarding of new members [6, 34] by outlining potential levers for action. It seems plausible that persistence and issue-related work can act as mentoring and thus as means to help newcomers get started [34]. For OSS practitioners, analyzing and possibly steering the observed relations may be helpful for increasing community size and thus potential human capital in their projects. Results also add to the discussion on the effects of balanced versus imbalanced configurations of subgroups [10, 11]. Our findings indicate that a larger share of members as an imbalanced configuration may have positive outcomes. Results further add to research on the differential effects of subgroups depending on their reason for formation and typology [11, 21].

7 Limitations and Future Research

This research is only a first step towards understanding the configurational properties of OSS members based on faultline and subgroup theory. There are several limitations, which in part may also explain the low pseudo R^2 values. A significant set of limitations arises from the choice of sampling and model specification. First of, the filtering criteria for including projects may have skewed results. In addition, the choice and operationalization of variables affect results. As faultline and subgroup measures, the entire breadth of characteristics studied in group research and psychology are conceivable. While carefully developed, operationalizations may not capture the phenomenon under study as expected. As an example, the operationalization of success as community size is just one option considering propositions to operationalize OSS success in multiple dimensions [29]. Furthermore, the inability to classify some comments may have skewed results. We strongly encourage further research using additional variables and testing the applicability of other operationalizations. Data has been collected during a limited timeframe using a lagged structure, which may have reduced explanatory power, especially if the timeframe studied was not representative for longer running projects. To inch closer to causal inferences, other methodologies such as experiments or mixed method approaches may be beneficial.

The distinction between core members and community poses two issues. Firstly, the threshold of attribution of members to either group may skew results. The operationalization may conflict with the observed fluid nature of OSS teams [23]. Secondly, we have studied relations of the share of members in the core team on the size and the extended community. This implies that all effects are indirect across the boundary of the core team, which may have caused some unexpected results.

Addressing these limitations and investigating additional aspects are promising avenues for future research. The effects of faultlines and subgrouping in virtual,

loosely coupled groups warrants further exploration. In addition, investigating the effects of subgroups on outside individuals seems promising. Moreover, it may be worthwhile to investigate interaction effects between the proposed subgroups. As an example, the share of members with a combination of high reputation and issue focus would further the investigation of the proposed onboarding mechanism provided by these factors.

8 Conclusion

Open Source development has become an established organizational way of building software. Performance effects of faultlines and subgroups are commonly discussed in team research. While most subgrouping is described to be detrimental for team outcomes, more recent works have proposed to consider different basis for subgrouping and their configuration. We have investigated the faultline-based subgroup concept in OSS projects by first identifying characteristics that may trigger faultlines and subgroups. This assertion is the basis for the empirical investigation of the relations of community size and the relative share of core members belonging to subgroups characterized by high levels of reputation, focus on issues, and extent as well as persistence of activity. We find significant relations with the size of the extended project community by contribution persistence and extent and positive albeit insignificant relations of a large share of members with high reputation and issue focus. Our results add to extant research on subgrouping and their configurational properties. In addition, they provide an additional step towards understanding how success as community size of OSS projects can be fostered.

9 Acknowledgements

This research and development project was funded by the German Federal Ministry of Education and Research (BMBF) within the Program Innovations for Tomorrows Production, Services, and Work (02K14A080) and managed by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the contents of this publication.

References

1. von Hippel, E., von Krogh, G.: Open Source Software and the “Private-Collective” Innovation Model: Issues for Organization Science. Organ. Sci. 14, 209–223 (2003).
2. Temizkan, O., Kumar, R.L.: Exploitation and exploration networks in open source software development: An artifact-level analysis. J. Manag. Inf. Syst. 32, 116–150 (2015).
3. Khondhu, J., Capiluppi, A., Stol, K.-J.: Is It All Lost? A Study of Inactive Open Source Projects. In: OSS 2013, IFIP AICT 404. pp. 61–79 (2013).
4. Hertel, G., Niedner, S., Herrmann, S.: Motivation of software developers in open source projects: An Internet-based survey of contributors to the Linux kernel. Res. Policy. 32, 1159–1177 (2003).

5. Stewart, K.J., Gosain, S.: The Impact Of Ideology on Effectiveness In Open Source Software Development Teams. Mis Q. 30, 291–314 (2006).
6. Steinmacher, I.. b, Graciotto Silva, M.A.., Gerosa, M.A.., Redmiles, D.F..: A systematic literature review on the barriers faced by newcomers to open source software projects. Inf. Softw. Technol. 59, 67–85 (2015).
7. Lau, D.C., Murnighan, J.K.: Demographic diversity and faultlines: The compositional dynamics of organizational groups. Acad. Manag. Rev. 23, 325–340 (1998).
8. Daniel, S., Agarwal, R., Stewart, K.J.: The Effects of Diversity in Global , Distributed Collectives: A Study of Open Source Project Success. Inf. Syst. Res. 24, 312–333 (2013).
9. Thatcher, S.M.B., Patel, P.C.: Group Faultlines. J. Manage. 38, 969–1009 (2012).
10. O'Leary, M.B., Mortensen, M.: Go (Con)figure: Subgroups, Imbalance, and Isolates in Geographically Dispersed Teams. Organ. Sci. 21, 115–131 (2010).
11. Carton, A.M., Cummings, J.N.: A theory of subgroups in work teams. Acad. Manag. Rev. 37, 441–470 (2012).
12. Howison, J., Crowston, K.: Collaboration Through Open Superposition: a Theory of the Open Source Way. MIS Q. 38, 29-A9 (2014).
13. Wiesche, M., Krcmar, H.: The relationship of personality models and development tasks in software engineering. In: Proceedings of the 52nd ACM conference on Computers and people research - SIGMIS-CPR '14. pp. 149–161 (2014).
14. Sahraoui, S., Al-Nahas, N., Suleiman, R.: An Agile Perspective on Open Source Software Engineering. Int. J. Open Source Softw. Process. 4, 52–65 (2012).
15. Crowston, K., Shamshurin, I.: Core-periphery communication and the success of free/libre open source software projects. J. Internet Serv. Appl. 8, (2017).
16. Tan, Y., Mookerjee, V., Singh, P.: Social capital, structural holes and team composition: Collaborative networks of the open source software community. In: ICIS 2007 Proceedings (2007).
17. Grewal, R., Lilien, G.L., Mallapragada, G.: Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems. Manage. Sci. 52, 1043–1056 (2006).
18. Jehn, K.A., Bezrukova, K.: The faultline activation process and the effects of activated faultlines on coalition formation, conflict, and group outcomes. Organ. Behav. Hum. Decis. Process. 112, 24–42 (2010).
19. Przybilla, L., Wiesche, M., Krcmar, H.: The Influence of Agile Practices on Performance in Software Engineering Teams: A Subgroup Perspective. In: Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research. pp. 33–40. ACM, New York, NY, USA (2018).
20. Faraj, S., Sproull, L.: Coordinating Expertise in Software Development Teams. Manage. Sci. 46, 1554–1568 (2000).
21. Carton, A.M., Cummings, J.N.: The impact of subgroup type and subgroup configurational properties on work team performance. J. Appl. Psychol. 98, 732–758 (2013).
22. Pflügler, C., Wiesche, M., Krcmar, H.: Subgroups in Agile and Traditional IT Project Teams. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018).
23. Vasilescu, B., Filkov, V., Serebrenik, A.: Perceptions of diversity on GitHub: a user survey. In: Proceedings of the Eighth International Workshop on Cooperative and Human Aspects of Software Engineering. pp. 50–56 (2015).
24. David, P.A., Shapiro, J.S.: Community-based production of open-source software: What do we know about the developers who participate? Inf. Econ. Policy. 20, 364–398 (2008).

25. Bezrukova, K., Jehn, K.A., Zanutto, E.L., Thatcher, S.M.B.: Do Workgroup Faultlines Help or Hurt? A Moderated Model of Faultlines, Team Identification, and Group Performance. Organ. Sci. 20, 35–50 (2009).
26. Bird, C., Gourley, A., Devanbu, P., Swaminathan, A., Hsu, G.: Open borders? Immigration in open source projects. In: 29th International Conference on Software Engineering Workshops(ICSEW'07) (2007).
27. Casaló, L. V., Cisneros, J., Flavián, C., Guinalíu, M.: Determinants of success in open source software networks. Ind. Manag. Data Syst. 109, 532–549 (2009).
28. Arazy, O., Daxenberger, J., Lifshitz-Assaf, H., Nov, O., Gurevych, I.: Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. Inf. Syst. Res. 27, 792–812 (2016).
29. Crowston, K., Annabi, H., Howison, J., Masango, C.: Towards a portfolio of FLOSS project success measures. In: Workshop on Open Source Software Engineering, 26th International Conference on Software Engineering (2004).
30. Meyer, B., Glenz, A., Antino, M., Rico, R.: Faultlines and Subgroups : A Meta-Review and Measurement Guide. Small Gr. Res. 45, 633 –670 (2014).
31. Antikainen, M., Aaltonen, T., Väisänen, J.: The role of trust in OSS communities - Case Linux Kernel community. In: IFIP International Federation for Information Processing. pp. 223–228 (2007).
32. Ridings, C.M., Gefen, D., Arinze, B.: Some antecedents and effects of trust in virtual communities. J. Strateg. Inf. Syst. 11, 271–295 (2002).
33. Herraiz, I., Robles, G., Amor, J.J., Romera, T., González Barahona, J.M.: The processes of joining in global distributed software projects. In: Proceedings of the 2006 international workshop on Global software development for the practitioner - GSD '06 (2006).
34. Fagerholm, F., Münch, J., Borenstein, J., Park, M.: The Role of Mentoring and Project Characteristics for Onboarding in Open Source Software Projects. In: ESEM'14 (2014).
35. Blincoe, K., Sheoran, J., Goggins, S., Petakovic, E., Damian, D.: Understanding the popular users: Following, affiliation influence and leadership on GitHub. Inf. Softw. Technol. 70, 30–39 (2016).
36. Ouellette, J.A., Wood, W.: Habit and intention in everyday life. Psychol. Bull. 124, 54–74 (1998).
37. Singh, P., Tan, Y., Mookerjee, V.: Network Effects: The Influence of Structural Capital on Open Source Project Success. MIS Q. 35, 813–829 (2011).
38. Subramaniam, C., Sen, R., Nelson, M.L.: Determinants of open source software project success: A longitudinal study. Decis. Support Syst. 46, 576–585 (2009).
39. Stewart, K.J., Ammeter, A.P., Maruping, L.M.: Impacts of license choice and organizational sponsorship on user interest and development activity in open source software projects. Inf. Syst. Res. 17, 126–144 (2006).
40. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994).
41. Bandura, A.: Social foundations of thought and action. Englewood Cliffs, NJ. (1986).