ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/271137551

Gender and Tenure Diversity in GitHub Teams

Conference Paper · April 2015

DOI: 10.1145/2702123.2702549

CITATIONS
90

READS
361

6 authors, including:

Bogdan Vasilescu
Carnegie Mellon University
66 PUBLICATIONS **1,300** CITATIONS

SEE PROFILE

Baishakhi Ray
University of California, Davis
20 PUBLICATIONS **849** CITATIONS

SEE PROFILE

M. G. J. van den Brand
Eindhoven University of Technology
224 PUBLICATIONS **2,611** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Modularity in (domain-specific) (meta)modeling languages View project

Project Continuous Integration View project

All content following this page was uploaded by Alexander Serebrenik on 20 January 2015.

The user has requested enhancement of the downloaded file.

Gender and Tenure Diversity in GitHub Teams

Bogdan Vasilescu[†§*], Daryl Posnett[†], Baishakhi Ray[†], Mark G.J. van den Brand[§], Alexander Serebrenik[§], Premkumar Devanbu[†], Vladimir Filkov[†*]
[†]University of California, Davis and [§]Eindhoven University of Technology
[*]vasilescu@ucdavis.edu, filkov@cs.ucdavis.edu

ABSTRACT

Software development is usually a collaborative venture. Open Source Software (OSS) projects are no exception; indeed, by design, the OSS approach can accommodate teams that are more open, geographically distributed, and dynamic than commercial teams. This, we find, leads to OSS teams that are quite diverse. Team diversity, predominantly in offline groups, is known to correlate with team output, mostly with positive effects. How about in OSS?

Using GITHUB, the largest publicly available collection of OSS projects, we studied how gender and tenure diversity relate to team productivity and turnover. Using regression modeling of GITHUB data and the results of a survey, we show that both gender and tenure diversity are positive and significant predictors of productivity, together explaining a sizable fraction of the data variability. These results can inform decision making on all levels, leading to better outcomes in recruiting and performance.

Author Keywords

Open source; gender; diversity; productivity; GitHub.

ACM Classification Keywords

H.5.3. [Information Interfaces and Presentation (e.g. HCI)]: Computer-supported cooperative work

INTRODUCTION

Because of the world-wide demand for talented and skilled labor, hiring in STEM (Science, Technology, Engineering, and Math) fields has become increasingly almost entirely meritocratic, and largely blind to demographic factors. This is certainly true for software engineering; as a result, both commercial and open source software teams can be very diverse. What are the effects of this on the project as a whole? Indeed, demographic similarity enhances mutual trust (and thus, arguably, team effectiveness), while demographic diversity may lead to stereotyping, cliquishness, and conflict [20, 43]. However, a team's social diversity seems to improve its technical performance [24].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2015, April 18 - 23, 2015, Seoul, Republic of Korea.
Copyright held by authors. Publication rights licensed to ACM. 978-1-4503-3145-6/15/04$15.00. http://dx.doi.org/10.1145/2702123.2702549

Software development teams can be *diverse* in various ways, *e.g.*, w.r.t. gender, experience, nationality, and coding language preference; some teams can be more diverse in one attribute and less so in others. Diversity attributes may also interact (*e.g.*, in some nations, female professionals may face more obstacles), which complicates analysis and study. Team diversity has been studied in physical ("meat-space") settings; however, data is hard-won in such settings. Smaller sample sizes make it difficult to effectively control for confounds. Data requirements for such effective controls, however, increase exponentially with the number of dimensions studied (one aspect of the "curse of dimensionality" [22]). Thus, studies of effects of diversity in teams (given the ineluctable confounds) require data on a great many teams, with sufficient variance along all co-variates of concern.

GITHUB, a social coding platform, has attracted millions of developers and thousands of Open Source Software projects.[1] All commits, issues, code changes, pull-requests etc. are archived and publicly available. GITHUB has become the new standard for comprehensive studies of social and technical organization and achievement [16, 37, 39, 41, 60]. Evidently, this is an attractive setting in which to study the relationship of diversity to performance. The scale of GITHUB is especially relevant when considering the role of women, who are very underrepresented in programming.[2] With a large enough dataset, however, the effect of increased gender diversity becomes noticeable. Additionally, since all data in GITHUB is historical (*i.e.*, archived), it is possible to study the effects of tenure, or one's length of time with a project and with GITHUB. However, the reliance on *volunteers* in OSS projects complicates matters; volunteers come and go, leading to team turn-over. Team turnover can certainly influence performance, and will confound the effects of diversity. The constructs of "team" and "team turnover" clearly also depend on the observation time-scale. In a healthy project, some rate of turnover is in fact desirable, as "new blood" brings in new abilities and ideas [21]. Arguably, turnover *will* affect observed diversity in GITHUB OSS teams, and must be considered carefully.

In this paper, using GITHUB data, we explore several questions: How diverse are online teams with respect to gender and tenure? Does gender diversity depend on tenure? On

[1]OSS depend on distributed volunteers' efforts whereas commercial software is much more centralized, and depends more on paid groups of programmers [23]; in both, the quality can be high [8].

[2]Especially so, it seems, in OSS projects: A 2013 FLOSS Survey [49] indicates 10% females; all earlier surveys [19] agree on merely 1–5%. Industry reports slightly higher numbers, *e.g.*, Google with 17% female technology employees.

team size? And how do gender and tenure diversity relate to outcomes like productivity? Remaining mindful of various confounds, we claim the following contributions.

(1) We collected a data set comprising thousands of projects from GITHUB, capturing or inferring the contributions, gender, and tenure of each recorded participant. We also ran a survey on GITHUB, to get a sense of how participants assess and value team composition and diversity.

(2) We use statistical modeling to analyze the relationship of gender and tenure diversity to productivity, when controlling for team size and other confounds. We find that *both gender and tenure diversity have a significant, positive effect on productivity*, gender across all team sizes and tenure for teams larger than 10. Together, these two explain 1–2.5% of the data variance, depending on team size.

(3) Models of turnover, or team change over time, in those teams reveal a negligible effect of gender diversity. Tenure has a large negative effect on turnover, while tenure diversity has a small, positive effect of turnover.

RELATED WORK

Background and Theory

The relationship between team diversity and team outcomes has been studied extensively over the past 40 years (see, *e.g.*, [31, 54] for recent reviews). Understood broadly, diversity arises from any attribute that differentiates people [66]. Attributes can be demographic (*e.g.*, age, gender, culture, ethnicity), functional (*e.g.*, role, tenure, expertise), or subjective (*e.g.*, personality). While there is ample literature on the matter (mostly from offline groups), different studies report different results. Some report significant positive correlations between diversity and performance (*e.g.*, [24]), while others report that diversity negatively impacts team outcomes (*e.g.*, [64]). Meta-analyses have tried to rationalize these contradictions by examining how effects are mediated by context (*e.g.*, group type, group size, or task complexity [31, 54]), research design, or sample characteristics [54].

However, despite the apparent contradictions between the different studies regarding the effects of diversity in work-groups, the literature tends to agree on three analytical frameworks: *similarity-attraction theory* (SA) [12], *social identity and social categorization theory* (SIC) [55], and *information-processing theory* (IP) [52]. According to SA, people prefer working with others similar to them in terms of values, beliefs, and attitudes [66]. SIC postulates that people tend to categorize themselves into specific groups, and categorize others as outsiders. Members of one's own group are then treated better than outsiders [55]. Due to greater perceived differences in values, norms, and communication styles between groups than within groups, SA and SIC explain why work-group heterogeneity can lead to confusion, stress, and conflict [31]. Both perspectives suggest negative effects of diversity on team outcomes. In contrast, IP treats diversity as positive: bringing to the table a mixture of cultural/educational backgrounds, and access to different networks and broader information can enhance a team's creativity, adaptability, and problem solving skills (*e.g.*, [33, 64]).

How team members react to diversity depends on the extent to which member characteristics are salient, and the effects are moderated by time [24, 30, 68]. As a team forms, members may well form durable first impressions of their new team-mates based on salient personal characteristics (typically demographic features), and as a result engage in social categorization [68]. However, as time passes and team members engage in meaningful interactions, the effects of surface-level diversity weaken and the team can better leverage its members' different cognitive frameworks and value sets (*i.e.*, their information-processing approaches), thus improving the team's efficiency and decision making processes [30]. Therefore, during a team's life-span, different phenomena related to the team's experience and size may arise, causing the effects of diversity to vary.

Diversity in OSS Teams

In task-oriented online communities (*e.g.*, OSS), group dynamics are substantially different than offline. First, in OSS geographic (and cultural) dispersion represents the norm, as members rarely meet in person; instead, collaborators on OSS projects assemble in online communities, and coordinate their activities through distributed communication channels (*e.g.*, email lists). Second, OSS teams are fluid, *i.e.*, they tend to form and dissolve organically around the task at hand. Third, since OSS contributors are often volunteers, teams face high turnover [50]. Finally, OSS communities generally exhibit a core-periphery structure [23], in which a small group of developers drives the work (the *core*), and a larger and more loosely coupled group of contributors (the *periphery*) supports the core by reporting issues, submitting patches, or contributing documentation. These characteristics make OSS teams less tangible than their offline counterparts: interactions between members are typically limited to online channels, and it can be hard to tell precisely who is on a team and who is not. Diversity attributes may, therefore, operate differently in OSS than in offline groups [40].

In particular, demographic characteristics such as ethnicity or gender become less salient in OSS [48]. Instead, OSS communities typically function as meritocracies [26], with sustained, high-quality contributions being the main drivers behind impression formation, reputation building, and trust [6, 16, 39]. For example, on social coding sites such as GITHUB, a very large fraction of members' actions within and across projects is more readily observable. This level of transparency, available without any personal interaction, allows one to make rich inferences about the commitment, work quality, or community significance of others [16, 17], and use these rather than demographics to form impressions.

The idea of OSS as meritocracy [26], widely-accepted, suggests that despite their great diversity (making them susceptible to SA or SIC), OSS teams should benefit from diversity. Diverse OSS teams should be able to reconcile and exploit their differences for the greater (*i.e.*, the project's) good, as IP suggests. However, all is not well in the meritocratic state of OSS. For example, the "hacker" culture tends to be male-dominated and unfriendly to women [58, p.194], who are underrepresented at not more than 10% of the all developers [19, 49]. Moreover, recent results[3] report active discrimination towards women in OSS, leading to the so-called "impostor syndrome": despite being knowledgeable

[3]FLOSSPOLS Deliberable D16 Gender, retrieved from `http://flosspols.org/deliverables.php` in Sep 2014

and professionally well-settled, women consider themselves to be disqualified or frauds. Moreover, sexist behavior in OSS is said to be "as constant as it is extreme" [44].

Very few authors have studied effects of diversity on team outcomes in online volunteer communities or OSS. Daniel *et al.* [18] studied effects of diversity on community engagement and market success in a sample of 357 SourceForge projects. They found that while diversity based on developers' reputation and role positively influences market success and community engagement, diversity of spoken language and nationality has a negative impact on community engagement, and a positive impact on market success. Chen *et al.* [13] studied diversity of experience and interests among members of Wikipedia Projects. They found that experience diversity increases group productivity and turnover up to a certain threshold, beyond which group productivity remains high but members are more likely to withdraw. In contrast, interest diversity increases productivity and decreases member withdrawal. In a similar context, Wang *et al.* [63] observed that members with very different tenure than the overall group tenure were more productive, but also more likely to withdraw early. To our knowledge, we are the first to consider effects of *gender* diversity in OSS communities.

Gender and Tenure Diversity in GitHub Teams

We choose to study gender and tenure diversity in GITHUB project communities. GITHUB (2008) has become the largest code host in the world, with many popular OSS projects (*e.g.*, the Linux kernel, Ruby on Rails, Bootstrap, Django, jQuery, and Homebrew) residing there. We chose GITHUB for our study because it: (i) is the largest ecosystem of its kind; (ii) has a wealth of trace data available [29]; and (iii) is known to promote women in technology through a number of initiatives, such as the Passion Projects, a series of talks aimed at providing female developers and engineers with role models, or the 2013 partnership with the Ada Initiative, offering free private repositories to women learning to write OSS. We are interested in effects of gender and tenure diversity on two project group outcomes: *productivity*, understood as the amount of work accomplished by each community of developers over time, and *turnover*, understood as the rate of change in each community over time.

Hypotheses

Gender is known to influence group outcomes [1, 42, 51] using the mechanisms explained by SA, SIC, and IP. Gender diversity brings different attitudes, perspectives, and values to a group. For example, while men are known to exhibit more task-oriented behaviors, women are said to show more socio-emotional behaviors [15], increasing a group's "ability to integrate divergent solutions while discouraging interpersonal and competitive conflict" [51]; this social orientation of women in teams is speculated to reduce the risk for egocentric listening and social loafing, and to allow for more effective teamwork [51]. As a consequence, diverse groups will be better equipped to handle uncertainty and take decisions, making them more productive (IP). On the flip side, gender diversity in a male-dominated "hacker" culture such as OSS may trigger discrimination towards women and sexism [44] (SIC), which in turn may lead to disengagement by women and perpetuation of the male dominance (SA).

The starting point for our exploration of gender diversity is the study of Rogelberg and Rumery [51]. The authors studied team decision quality, time on task, and interpersonal cohesion for 96 four-person teams having all five possible gender ratios (from all-male to all-female). The task, a winter survival exercise where each team had to rank order the most important 12 items salvaged from a plane crash, was determined to be male-oriented after pilot testing. The authors found that gender diversity significantly affected team decision quality (but not time on task nor interpersonal cohesion), with lone-female teams (*i.e.*, three men and one woman) performing the best. That is, despite the male orientation of the task, all-male teams were outperformed by the more gender-diverse lone-female teams. Apesteguia *et al.* [1] reported similar findings in a recent study of three-person teams competing in an online business simulation game: again the more gender-diverse teams (two men, one woman) performed the best. In both studies teams consisted of students. Since neither studies considered larger teams, it is hard to reason whether increasing gender diversity or rather simply having women on teams would be responsible, in general, for the reportedly positive effects on team outcomes. Given the previous argument based on IP, and supported by the parallel between the male-orientation of the task of Rogelberg and Rumery [51] and the known male-dominance of OSS, we posit, more generally, that:

H1. Gender diversity has a positive effect on productivity.

The socio-emotional behavior and non-aggressive strategies attributed to women in teams [42, 51] could imply a more positive attitude towards teamwork and collaboration in gender diverse projects. As all male-groups tend to be overaggressive and competitive [2], the presence of women may moderate over-competitiveness, reducing conflicts and improving communication and collaboration [51]. All other things aside, projects with a more welcoming community should exhibit less turnover, or:

H2. Gender diversity has a negative effect on turnover.

Tenure refers to the amount of experience accumulated by each group member. Highly tenured developers are more likely to be part of a project's core team (*i.e.*, to have committer status, the highest trust level in a typical OSS project), and to take part in decision making. Newcomers, or less tenured developers, are at the other end of the scale. Although typically not as experienced, they are very motivated to provide contributions by reputation building [36] and their desire to learn [67]. Mixed-tenure OSS communities will have more opportunities to leverage the energy and experience brought forward by juniors and seniors, respectively [13, 18], and will be more productive. The presence of tenured, senior developers will set the example for newcomers that increasing rewards can be obtained through sustained participation and contribution [18], and will create mentoring opportunities [47]. In turn, motivated by example [16, 28], newcomers may expend more efforts to prove themselves and build reputation [36]. Following this logic we test:

H3. Tenure diversity has a positive effect on productivity.

Tenure diversity tends to be associated with differences in attitudes, views, or approaches to tackle a particular problem [18]. For example, tenured developers may have a preference for more mature technologies, while newcomers may be more willing to experiment with newer, less proven, libraries or APIs. In the extreme, as tenure diversity in a project team increases, newcomers and seniors may find insufficient common ground, leading to communication breakdown and conflict [13], which may lead to higher attrition [46]. Since tenure-diversity-based conflicts have already been reported for Wikipedia Projects [13], a community comparable to OSS [45], we test whether:

H4. Tenure diversity has a positive effect on turnover.

METHODS AND DATA

We followed a mixed-methods approach with a sequential exploratory strategy [25]. First, we conducted a survey of GITHUB contributors to obtain insights in the way team collaboration and diversity are being perceived. Following it, to assist in interpreting the qualitative findings, we assembled a longitudinal dataset of GITHUB developers and projects, on which we modeled productivity and turnover outcomes using a variety of statistical models.

GITHUB's pull-based development model imposes a different definition for *project* than in traditional OSS. On GITHUB, to facilitate scaling up, typically only a small group of developers (*the core*) have access to the main repository and can push their changes directly. All other contributors, if any, work in isolation in their local copies of the repository (*forks*), and submit their contributions back for review and integration via *pull requests*. Therefore, a project on GITHUB consists of a base repository and all its forks (and transitively all their forks). We adopt this definition since measuring the activity of a repository independently of its forks will underrepresent the activity of all of them as part of a single project [34].

Survey

We selected 4,500 GITHUB contributors stratified according to gender (as inferred by `genderComputer` [59], *i.e.*, male, female, or unknown; see *Inferring gender* below) and number of projects contributed to (one project—the majority of contributors; many projects—7 or more, distributional outliers; and few projects—between 2 and 6), randomly selecting for each combination of gender and number of projects 500 individuals with known email addresses in our GHTorrent-based data set. We contacted all of them by email and invited them to participate in the survey. Participation was voluntary and confidential, and was expected to take about ten minutes. We received 816 responses to the survey and 236 automatic replies. Ignoring the latter, we achieved higher than 19% response rate, slightly exceeding the 15–17% response rates reported in comparable studies [4, 57].

Among the respondents, 199 (24%) indicated their gender as "female", an underrepresentation as at least one third of the invitations had been sent to women ("female" inferred by `genderComputer`). Age ranged from 14 to 66 (median 29; mean 30); IT experience ranged from zero (*e.g.*, a student, a manager, and an individual not working in tech) to 44 years (median 8; mean 10.5). The largest group of respondents reside in the USA (264), followed by Germany (52), France (42) and Canada (34). To evaluate the representativeness of the respondents group, we aggregate the country information to the macro-regional level (Africa, Asia, Australia and New Zealand, Eastern and Southern Europe, Latin America, North America, Western and Northern Europe). A χ^2 test comparing the macro-regional distribution of our respondents with those reported in a previous study of GITHUB users [56] reveals no differences w.r.t. the known general GITHUB population ($p > 0.9$).

Our first observation after analyzing the responses relates to perceptions of teams. Two-thirds of respondents indicated they consider themselves as part of a team when working on a repository. Women (76%) score higher than men (63%), in adherence with their more socio-emotional behavior in teams [42, 51]. When asked who else they consider part of their team, respondents could indicate multiple answers related to the kind of activity (*e.g.*, committing or issue reporting), frequency, and organizational aspects. The most popular answer was "everyone who does something in this repository (*e.g.*, pushes code, submits pull requests, reports issues)" (44%). Differences between answers by men and women were not statistically significant (χ^2 $p \simeq 0.97$).

Another observation relates to perceptions of diversity attributes. We asked whether they are aware of certain aspects of other team members (*e.g.*, age, gender, hobbies, political views, programming skill, real name, and social skills). 98% of those that answered this question indicated they are aware of the *programming skills* of at least some team members. This makes programming skills the most visible diversity aspect we considered, hardly surprising as GITHUB contributors can be expected to interact in their software development roles. In addition, almost half of respondents to this question (48.6%) indicated they are aware of the *gender* of most of their teammates. This salience of gender among team members on GITHUB contradicts earlier claims of obscurity of gender in OSS [48] and goes against the purely meritocratic model of OSS. Discrimination and sexism in OSS *can* happen if gender is so salient, and they *do* happen, as discussed in the previous section. This suggests that gender diversity in GITHUB teams may have more intricate effects than hypothesised above. Finally, as expected, hobbies and political views (3% are aware for most team mates; 17%—for some team mates) are the least visible diversity aspects.

Lastly, respondents greatly differed in perception of importance of diversity. Some maintain that OSS development is "more about the contributions to the code than the 'characteristics' of the person" (40, male, N America), and that "any demographic identity is irrelevant" (26, female, N America) since "code sees no color or gender" (34, male, W&N Europe). In contrast, *e.g.*, awareness of the only woman on the team of other team members' views on diversity makes her "feel more welcome and eager to work since the stereotype threat isn't prevalent" (29, female, W&N Europe), while diversity is being perceived as a "source of creativity" (42, male, N America). Unfortunately, not all diversity-related experiences are positive, *e.g.*, forcing a 23-year old N-American female respondent to create a fake GITHUB-handle to masquerade as male or to quit a project; the latter experience has also been reported by a 40-year old N-American female. A 38-year old N-American female summarized her experiences as mostly positive, in particular on projects she leads; for other projects "interactions are usually positive too, with occasional sexism, but nothing more than one encounters in the rest of life".

Overall, we draw several conclusions from the survey and carry these into the subsequent analysis. First, *participants recognize team-members as those who make any type of contribution to the project*. Second, *team members are quite aware of certain aspects of other team members, including gender and technical skill*. Finally, *diversity appears to matter to contributors, but the perceived effects appear to vary*.

Data Set

Assembling a *diversity* GITHUB data set, with precise gender, location and tenure information, is challenging. First, on a GITHUB profile page personal information fields (*e.g.*, name, location) are free-text entries, thus often noisy; gender data is missing altogether. Second, some developers use multiple accounts, making it difficult to accurately measure their contributions. Finally, the OSS population, with the typically heavy-tailed workload distributions [62] (few very active projects/contributors and many much less active ones) require careful data filtering and preprocessing.

Preprocessing and Filtering

Our data collection process starts from the MySQL GHTorrent dump [29] dated 1/2/2014. GHTorrent is an offline persistent mirror of GITHUB's event streams, containing all user events, including commits, comments, pull requests, etc.

We then cleaned and filtered this data set as follows: (i) we unified different accounts used by the same contributor (see unmasking aliases below); (ii) we only considered activity between January 1, 2008 00:00 and January 2, 2014 23:59; (iii) we removed inactive projects (*i.e.*, having strictly less than 100 commits in total, and strictly less than 90 days of history between their earliest and latest recorded events, be it commits, pull requests, issues, or comments); (iv) since we are interested in effects of diversity on *team* outcomes, we excluded very small projects (*i.e.*, having less than four contributors, two of which committers, throughout their history); and (v) since we could not infer gender for all contributors (see the following subsection) and in order to compute gender diversity reliably, we further excluded projects for which we could not infer gender for at least 75% of contributors. Table 1 presents basic statistics about our data set. The effects of steps (ii)–(v) are denoted by "post filter". We stress the drastic but necessary reduction in number of projects during filtering. Even though our sample is restricted to only 1% of the initial projects, this step is paramount to ensuring sufficient variance in the data set along all co-variates of concern.

Unmasking Aliases

Recognising all the different accounts (or aliases) a person may have used when contributing to an online community (process known as identity merging or unmasking aliases) is essential to attributing contributions accurately, yet an ever-recurring challenge when analyzing OSS trace data [27, 35, 61]. On GITHUB, the committer's name and email address are set locally in each developer's git client, rather than globally at GITHUB level. Hence, due to variations in these attributes across devices or time, commits

Attribute	Total	Post filter	Min	Med	Mean	Max
Repo's	6,818,042					
Projects	2,605,486	23,016				
Commits	76,478,208	10,735,308	100	211	466.4	150,380
Pull req's		1,827,391	0	11	79.4	135,557
Issues		697,913	0	4	30.3	22,905
Comments		2,633,614	0	11	114.4	102,813
Contrib's	2,677,443	671,301	4	10	29.1	13,230
Females		9,292	0	0	0.96	300
Males		157,965	0	8	22.9	9650

Table 1. Basic statistics about our GITHUB data set.

may not be attributed properly to the relevant GITHUB accounts. Furthermore, to account for "unknown" committer aliases encountered while processing events from GITHUB's API without losing data, GHTorrent assigns the corresponding commits to artificial aliases not linked to actual GITHUB accounts. This phenomenon is far from negligible: in the GHTorrent dump we analyzed, approximately 23% of all users have been labeled as "unknown".

To link the different aliases belonging to the same GITHUB contributors as well as deal with the issue of "unknown" aliases (an artifact of GHTorrent), we devised a series of heuristics inspired by those of Bird *et al.* [5] and Vasilescu *et al.* [62].[4] We stress that we have been as conservative as possible when deciding if two aliases belong to the same person, as the false positives count (*i.e.*, aliases incorrectly merged) would otherwise likely increase significantly [27]. For example, if multiple aliases share the same well-formed and non-fictitious email address, then we merge them, since email addresses are considered individual. Otherwise, if email addresses differ, then we only merge aliases for which we have collected sufficient evidence that they belong to the same person, from their first and last names, usernames, email address prefixes and domains, or locations. Even so, with a conservative approach to unmasking aliases, we determined that more than 170,000 users in our data set had used more than one alias (2,917,942 before; 2,677,443 after, 91.7%), with a median of 2, a mean of 2.4, and a maximum of 14. Moreover, using these heuristics we linked more than half of the "unknown" users to actual GITHUB accounts. Each unique account after unmasking aliases contained activity and personal data aggregated from all its aliases.

Inferring Gender

For each GITHUB contributor we infer gender based on their name and, if available, country, following the approach of Vasilescu *et al.* [59]. This approach combines a number of transformations, diminutive resolution, and heuristics (*e.g.*, users from Russia with surnames ending in *-ova* are female), with female/male frequency name lists collected for thirty different countries. Country data is indeed an essential ingredient to inferring a person's gender from their name (the classical example is Andrea, a common male first name in Italy, but a common female one in many other countries). In absence of country information, this gender inference approach seeks agreement between all country name lists available that report on a given name. The reported precision of `genderComputer` is 93% [59].

[4] Our code, too involved to be presented here, is available online at **https://github.com/bvasiles/ght_unmasking_aliases**.

On GITHUB, registered users can choose to disclose their name and location in their profiles, but these fields are unstructured and often noisy, if not absent altogether. For example, only 40% of all users identified after unmasking aliases filled in what we consider reliable entries (*i.e.*, at least two parts separated by space) in their name fields. Location descriptions are also not readily trustworthy. Since not all include actual geographic data (*e.g.*, city names, latitudes/longitudes, or postcodes), to identify countries we combine information obtained from the Bing Maps API with information derived using our custom set of heuristics.[5] Following this process we identified countries for 12.6% of all users. Then, using Vasilescu *et al.*'s `genderComputer` [59], we could infer gender for 873,392 (32.6%) GITHUB contributors: 91% male, 9% female. These numbers confirm an underrepresentation of women on GITHUB, but comparable proportions to those reported for OSS (*e.g.*, [19, 49]) or Stack Overflow [59].

We stress that while the fraction of users for which we could identify gender seems low (32.6%), it represents 80% of those who disclosed their names, the best we could hope for using any name-based technique, be it automatic or manual. Among the remaining 20% without any gender inferred, manual inspection of a small sample revealed missing country data as the main cause of ambiguity.

Measures

Using the methodology described above we assembled a longitudinal dataset of GITHUB projects, in which each observation contains the composition, characteristics, and outcomes of a project's team of contributors for each quarter in the evolution of the project. We measure quarters as 90-day intervals starting from the earliest recorded commit in each project's history, such that data for the last quarter included is not incomplete, *i.e.*, there is at least one commit on the end date of this last quarter or later. In line with the study of diversity in Wikipedia projects of Chen, Ren, and Riedl [13], we also stress that a longitudinal approach is crucial: diversity may affect turnover, which in turn affects team composition, which affects diversity.

Response variables

Productivity: We measure team productivity by the number of commits by team developers recorded in either the main repository or any of its forks in a given quarter. Commits are the most encompassing form of coding contribution to a GITHUB project and a representative facet of developer productivity in OSS [18]. Clearly, not all commits are created equal, some have different length and quality, and only some commits to forks are integrated in the main repository. Therefore, their number is insufficient to quantify the total contribution of a team in a project, or even to quantify their energy expenditure while doing so. However, commits are a reasonable representative, or sample, of the overall activities developers undertake while working on a project, hence a reasonable estimate of team productivity.

Turnover: As discussed earlier and as supported by our user survey results, we define each project's team as the set of contributors active in a given quarter, where activities can be either commits, pull requests (merged or not), issues, or comments, recorded in the project's main repository or any of its forks. To capture turnover in project teams we use the fraction of the team in a given quarter that is different with respect to previous quarter (*i.e.*, the turnover ratio).

Independent variables

Gender diversity: We measure team gender diversity using the Blau index [10], defined as $1-\sum_{i\in\{m,f\}} p_i^2$, where p_i are the fraction of male and female team members. The Blau index is a well-established diversity measure for categorical variables (*e.g.*, [13, 33]). If teams contain members for which we could not infer gender, the measure only considers the fraction of the team for which we could infer gender.

Tenure diversity: Different kinds of tenure are accessible for GITHUB developers: (i) account tenure, capturing global GITHUB *presence*; (ii) commit tenure, capturing global GITHUB *coding experience*; or (iii) project tenure, capturing local *project experience*, not restricted to coding. We select commit and project tenure as the most appropriate measures of coding experience and of project expertise, respectively. Note that while GITHUB commit tenure is restricted to coding experience *within* GITHUB, other, perhaps more appropriate, measures of coding experience *outside* GITHUB are unavailable. To measure commit tenure for a GITHUB contributor w.r.t. a given quarter, we compute the number of days since her earliest ever recorded commit (in any GITHUB repository) until the end of that quarter. Commit tenure is not available for all team members, since not everyone will have been committer, in this or other projects. To measure project tenure for a GITHUB contributor, we compute the number of quarters since her earliest recorded event in the current project until the end of that quarter. Project tenure is available for all team members. Commit and project tenure are orthogonal measures: a team can be diverse w.r.t. commit tenure, if it has a mix of novice and expert coders, but homogenous w.r.t. project tenure, if all contributors formed the team at the same time. Since tenure measures are numerical variables, we measure tenure diversity using the coefficient of variation, defined as the ratio between the adjusted standard deviation and the mean.

Control variables

Team size: The number of contributors per project team in a given quarter. Larger team sizes are likely associated with increased productivity and increased turnover. However, given a small overall number of women involved in OSS projects, we expect the impact of gender diversity to be more visible for smaller teams.

Project forks: The number of forks part of a project. Since activity in forks is more likely to be limited (both in time and in amount) compared to a project's main repository, the number of forks is expected to have a negative effect on team productivity, and a positive effect on turnover.

Quarter index: The index of each 90-day interval in a project's history, starting with 1. Time is known to have a moderating influence on the effects of diversity [64]: as group members continue to interact, they have more time to adjust to differences between them.

Overall project activity: The overal total commit count.

[5] `https://github.com/tue-mdse/countryNameManager`

Project age: The difference between the maximum index and the index of the 90-day interval during which the first commit was recorded, w.r.t. 1/1/2008 (newer projects have smaller values), starting with 24. Controls for changes in environment as GITHUB grows with time: later projects and their teams may have experienced a different culture.

Tenure median: To represent the "average" project or commit tenure as opposed to tenure diversity, we also include the project median tenure and the commit median tenure.

Comments: The number of GITHUB comments left by the team members during a given quarter on commits, pull requests, or issues. Reflects a project's social activity (not covered by our productivity measure).

Statistical Modeling

We wish to capture the variation in several outcomes with respect to the basic control measures and diversity indices described above. Our dataset contains a large number of projects and, within each project, consists of multiple time windows. Mixed effects models are often used in this context in order to capture measurements from within the same group, *e.g.*, within the same project, as a random effect. Similarly, they are used longitudinally, to measure the variance between time points. In our case we have both, so we use two random effect terms, the project and time window in which the measurement was taken. For the time window we model only the intercept as a random effect. However, as we do not expect that projects will respond to different team sizes in the same way, *e.g.*, some types of projects may benefit greatly from the addition of a new team member, while others less so, we include team size as a random slope within the project. All other variables were modeled as fixed effects. We tested the inclusion of random team size slope within time windows but the additional random effect was not significant. We used multiple linear mixed-effects models, as implemented in the functions *lmer* and *lmer.test* from the *lme4* [3] package in R.

Due to the varying (*i.e.*, unstable) behavior of the fitted coefficients in the models over the full range of team sizes, we decided it was appropriate to use *piecewise*, or *segmented* regression models. In those models, the range of a variable is split at structural breakpoints and multiple, piecewise linear models are fitted, one for each segment. Additionally, as we are interested in increased effects of diversity on OSS outcomes, we further required at least one female committer over the life of the project. This better corresponds to our hypothesis of increasing diversity, when some diversity is already present, versus adding a woman to an all-male team. Coefficients are considered important if they were statistically significant ($p < 0.05$). Their effect sizes are obtained from ANOVA analyses. We evaluate models using the sum of squares goodness of fit measure, while insisting on coefficient stability across models, in the segmented regression.

RESULTS

To assess the effects of our controls and diversity predictor variables on productivity and turnover we used mixed effects, multiple linear regression models. We filtered the data and retained a subset of it that both allowed for the calculation of all measures and was homogenous enough to yield stable models, as described in the Data section above. The variables that upon examination showed a log-normal distribution (*e.g.*, counts) were log-transformed, as indicated. A number of different measures that were highly correlated were excluded from the models. We note that our models were driven by our hypotheses, and were all based on the full complement of variables and confounds that we started with, concordant with good statistical modeling practice [53].

Productivity

To model productivity, we used *segmented regression* due to the instability of the fitted model coefficients when a single model was used over the whole range of team sizes. The results of three models, one for small teams (< 11), the second for medium-sized teams (11 to 30 members), and a third, for large teams (> 30), are given in Table 2. In addition to the three model coefficients, given in the left columns of the three pairs of model columns, the table shows the sum of squares, a measure of variance explained, for each variable. The significance is indicated by stars. We note that the project coefficients and the sum of squares are not comparable across the three models, but only within each model.

As expected, the number of committers and the other controls play a dominant role in explaining the variance in the data. The effect of gender diversity is positive, highly significant, and stable over all team size ranges. Its contribution to explaining the data variance is sizable for those team sizes. Thus, ***H1*** is confirmed. Likewise, looking at the commit tenure diversity coefficient, we see it is positive and significant for medium and large teams, partially confirming ***H2***.

Turnover

The turnover models are given in Table 3. We did not observe an effect instability with respect to team size in the turnover models, hence we did not use segmented regression. As with the productivity models, the number of committers and the other controls play a dominant role in explaining the variance in the data. For turnover, gender diversity is not significant in the model, thus we could not confirm a relationship between increasing gender diversity and the tendency for team composition to change. Therefore, ***H3*** is not confirmed. Both tenure diversity measures, however, are significant and contribute significantly and positively to the variance explained by the model. Therefore, ***H4*** is confirmed. Additionally, the interaction with median tenure across the team is also significant; we discuss this further in Section 4.

DISCUSSION

In the first set of models, we find that gender and tenure diversity have very significant, positive effects on productivity, across different team size segments, when controlled for other effects. These effect sizes are small compared to the controls: together, the two explain between 1% and 2.5% of the total sum of squares (variance). This falls into the range of small, but reportable effects (0.01 or greater in Cohen's criteria [14], which admittedly are not without fault [38]). Their small sizes were expected since the majority of the variance is explained by the big, technical players: total commits, project age, team size and number of committers. The effects, however, are stable (remain significant) across different team sizes. A plausible interpretation of our results is, then, that given strong known technical confounds, there is

	Small Teams		Medium Teams		Large Teams	
	Coeffs (Errors)	Sum Sq.	Coeffs (Errors)	Sum Sq.	Coeffs (Errors)	Sum Sq.
(Intercept)	−0.16281 (0.04003)***		0.05540 (0.02247)*		0.10205 (0.02667)***	
scale(log(total_commits))	0.82962 (0.00772)***	3127.67***	0.92775 (0.01961)***	487.19***	0.84236 (0.03348)***	122.99***
scale(log(proj_age + 0.5))	−0.08517 (0.00760)***	696.09***	−0.15728 (0.01785)***	94.50***	−0.13490 (0.02363)***	26.06***
scale(log(num_team))	0.34781 (0.00742)***	388.06***	0.13788 (0.01216)***	14.09***	0.24916 (0.03403)***	6.01***
scale(log(total_committers))	−0.35371 (0.01068)***	874.25***	−0.16661 (0.02288)***	37.08***	−0.07542 (0.04020)	4.68
scale(log(forks + 0.5))	−0.11534 (0.00915)***	33.53***	−0.16243 (0.01951)***	9.75***	−0.14167 (0.03411)***	1.78***
scale(log(num_comments + 0.5))	0.06840 (0.00626)***	24.18***	0.08385 (0.01588)***	3.85***	0.07763 (0.02582)**	1.47**
scale(proj_tenure_med)	0.05559 (0.00770)***	24.64***	0.00500 (0.01651)	0.30	0.02607 (0.02667)	0.67
scale(proj_tenure_div)	−0.04899 (0.01126)***	7.37***	−0.10813 (0.01940)***	6.49***	−0.09226 (0.02208)***	1.97***
scale(commit_tenure_med)	−0.02622 (0.00754)***	12.80***	0.05927 (0.02100)**	0.00**	0.14005 (0.03104)***	0.50***
scale(commit_tenure_div)	0.01395 (0.00750)	0.53	0.07525 (0.02195)***	5.17***	0.12343 (0.03316)***	2.18***
scale(gender_div)	0.03137 (0.00564)***	8.59***	0.03765 (0.01043)***	2.20***	0.03992 (0.01522)**	0.84**
scale(proj_tenure_med):scale(proj_tenure_div)	0.01554 (0.00933)	1.00	0.02166 (0.02226)	0.16	0.05869 (0.02828)*	0.72*
scale(commit_tenure_med):scale(commit_tenure_div)	0.00622 (0.00498)	0.43	0.00045 (0.00966)	0.00	0.01893 (0.01190)	0.29

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2. Productivity Models, log (number of commits) is the response. Team size segments: Small < 11; $10 <$ Medium < 30; $30 <$ Large.

	Diversity	
	Coeffs (Errors)	Sum Sq.
(Intercept)	−0.26408 (0.07017)***	
scale(log(total_committers))	0.32924 (0.01120)***	708.47***
scale(log(forks + 0.5))	0.04446 (0.00883)***	244.04***
scale((proj_age + 0.5))	−0.00526 (0.00788)	49.17
scale(log(total_commits))	−0.44783 (0.00819)***	2415.86***
scale(log(num_team))	0.19782 (0.00753)***	773.27***
scale(commit_tenure_med)	0.02439 (0.00741)**	92.86**
scale(commit_tenure_div)	0.07222 (0.00741)***	274.39***
scale(proj_tenure_med)	−0.21963 (0.00758)***	933.91***
scale(proj_tenure_div)	0.28548 (0.01191)***	169.13***
scale(gender_div)	0.00756 (0.00528)	2.25
scale(log(total_committers)):scale(log(forks + 0.5))	0.01607 (0.00497)**	4.10**
scale(commit_tenure_med):scale(commit_tenure_div)	−0.04586 (0.00465)***	24.12***
scale(proj_tenure_med):scale(proj_tenure_div)	0.09921 (0.00910)***	42.80***

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3. Turnover Models, developer turnover is the response.

a small but notable effect of gender and tenure diversity on productivity in online software teams.

The relative effects of gender and tenure vary with team size. As gender diversity increases, team productivity increases. The coefficient and the size of the effect are roughly the same across the three separate team size ranges. The positive effect is not surprising in light of prior work on online [13] and offline groups [24]. Our findings align more with IP [52] than the other theories discussed in the background section; this suggests that merit matters more than demographics in OSS. In general, in OSS projects the existence of latent social structures (communities) corresponding to code modules has been documented [7]. Thus, project participants are arguably acutely aware of others in their smaller community, and are more attuned to reacting to changes within those smaller groups. On average, our results across team sizes point to the equitable distribution of gender diversity across the communities, over time, resulting in a scalable diversity effect, which we observe. The productivity of programming teams as a function of size is a complicated issue [11, 65]; however, one can reasonably expect that the effects of size in this data are modulated by software modularization and the organization of teams into task-oriented groups.

We now consider diversity in commit tenure and project tenure. Recall that commit tenure measures experience across *all* of GITHUB, not just the specific project under consideration; thus, it reflects the experience diversity of the individuals in the team. On the other hand, project tenure diversity reflects the diversity of experience of project members *within the project*. Thus, it is entirely possible for a project to be highly commit-tenure diverse and not very project-tenure diverse (and vice-versa). The effects of both types of diversity tend to grow with team sizes. Commit-tenure diversity has no significant productivity effect for smaller teams. This is consistent with the theory that introducing "new blood" benefits any project, and more so the bigger ones. In contrast, project tenure diversity is negative and significant, indicating that a lot of per project diversity is counterproductive. This suggests that project diversity increases co-ordination effort, perhaps arising from conflicts between vested interests (long tenure) and newcomers.

The second model shows that turnover (rate of change in project staff) is not effected by gender diversity, since its coefficient is not significant. So, higher diversity does not effect a change in team composition. These models do show significant effects of tenure on turnover, however. In particular, median project tenure (not tenure diversity) has a substantial negative effect on staffing changes, *i.e.*, the higher the median tenure in the project, the less likely it is there will be people joining or leaving the project. This is consistent with the notion that more experienced teams will persist, and attrition will be lower. The project and commit tenure diversity effects are both positive: more diverse teams experience more staffing changes. This is due to, perhaps, people preferring teams with more homogeneous experience.

THREATS TO VALIDITY

Any empirical study is subject to a number of threats to validity. Validity of extraction of information from GITHUB can be threatened in a way similar to digital trace data [32] or put in a "peril" by concerns related to mining Git repositories in general [9], or specifically GITHUB data [34].

We attended to the threats identified above; *e.g.*, our "post filter" steps discussed earlier removes projects with too few commits (Peril II [34]) and also one-person projects (Peril V [34]). The threat of misrepresenting individual contributions, related to generated-data reliability [32], was addressed by combining contibutions of the different aliases of an individual, and in forks of the same project. However, despite best efforts, our measures of productivity (number of commits) might be an underestimation, since the project history might have been rewritten (Perils 4 and 7 [9]). Furthemore, to reduce the threats related to analysis tools [32] we have opted, whenever possible, for techniques and approaches empirically evaluated in previous studies: *e.g.*, the gender inference tool has been evaluated previously [59], the Blau index is a common diversity measure [13,33], and temporal aggregation, concern raised before [32], is done per quarter as common in diversity studies [13,18,63].

CONCLUSION

Our study suggests that, overall, when forming or recruiting a software team, increased gender and tenure diversity are associated with greater productivity. This finding is consistent with information processing theory [52]: a diverse team, comprised of women and men, and more and less experienced people, brings different perspectives together, thus improving outcomes. The benefits of experience diversity have a limit, though, as higher tenure diversity may increase attrition. This negative effect appears to be mitigated, however, when more experienced people are present.

Currently, however, women programmers are in the minority in OSS and other technical teams. Our study suggests that on a larger, economic and societal scale, added investments in educational and professional training efforts and outreach for female programmers will likely result in added overall value. Such efforts may have to be coordinated with efforts to remove existing barriers standing in the way of women's equitable integration in programming teams. To the best of our knowledge this is the first work that provides empirical evidence that productivity and turnover of software development teams are effected by team diversity.

ACKNOWLEDGEMENTS

Many thanks to survey participants for their time and interest, and to reviewers for their helpful remarks. BV, AS, and MvdB gratefully acknowledge support from the Dutch Science Foundation, project NWO 600.065.120.10N235. VF, PD, DP, BR, and BV gratefully acknowledge support from the US Air Force Office of Scientific Research, project FA955-11-1-0246.

REFERENCES

1. Apesteguia, J., Azmat, G., and Iriberri, N. The impact of gender composition on team performance and decision making: Evidence from the field. *Manage. Sci. 58*, 1 (2012), 78–93.
2. Aries, E. Interaction patterns and themes of male, female, and mixed groups. *Small Group Behavior* (1976).
3. Bates, D. M. *lme4: Mixed-effects modeling with R*, 2010.
4. Begel, A., and Nagappan, N. Usage and perceptions of agile software development in an industrial context: An exploratory study. In *ESEM*, IEEE (2007), 255–264.
5. Bird, C., et al. Mining email social networks. In *MSR*, ACM (2006), 137–143.
6. Bird, C., et al. Open borders? immigration in open source projects. In *MSR*, IEEE (2007), 6.
7. Bird, C., et al. Latent social structure in open source projects. In *FSE*, ACM (2008), 24–35.
8. Bird, C., et al. Does distributed development affect software quality?: An empirical case study of Windows Vista. *CACM 52*, 8 (2009), 85–93.
9. Bird, C., et al. The promises and perils of mining git. In *MSR*, IEEE (2009), 1–10.
10. Blau, P. M. *Inequality and heterogeneity: A primitive theory of social structure*, vol. 7. Free Press NY, 1977.
11. Brooks, F. P. *The mythical man-month*, vol. 1995. Addison-Wesley Reading, MA, 1975.
12. Byrne, D. E. *The attraction paradigm*. Personality and psychopathology. Academic Press, 1971.
13. Chen, J., et al. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *CHI*, ACM (2010), 821–830.
14. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.
15. Croson, R., and Gneezy, U. Gender differences in preferences. *J. Econ. Lit.* (2009), 448–474.
16. Dabbish, L., et al. Social coding in GitHub: transparency and collaboration in an open software repository. In *CSCW*, ACM (2012), 1277–1286.
17. Dabbish, L., et al. Leveraging transparency. *IEEE Software 30*, 1 (2013), 37–43.
18. Daniel, S., et al. The effects of diversity in global, distributed collectives: A study of open source project success. *Inform. Syst. Res. 24*, 2 (2013), 312–333.
19. David, P. A., and Shapiro, J. S. Community-based production of open-source software: What do we know about the developers who participate? *Information Economics and Policy 20*, 4 (2008), 364–398.
20. de Gilder, D., and Wilke, H. A. M. Expectation states theory and the motivational determinants of social influence. *Eur. Rev. Soc. Psychol. 5*, 1 (1994), 243–269.
21. de O. Melo, C., et al. Interpretative case studies on agile team productivity and management. *Inform. Software Tech. 55*, 2 (2013), 412–427.
22. Domingos, P. A few useful things to know about machine learning. *CACM 55*, 10 (2012), 78–87.
23. Ducheneaut, N. Socialization in an open source software community: A socio-technical analysis. *CSCW 14*, 4 (2005), 323–368.
24. Earley, C. P., and Mosakowski, E. Creating hybrid team cultures: An empirical test of transnational team functioning. *Acad. Manag. J. 43*, 1 (2000), 26–49.
25. Easterbrook, S., et al. Selecting empirical methods for software engineering research. In *Guide Adv. Empir. Softw. Eng.* Springer, 2008, 285–311.
26. Feller, J., and Fitzgerald, B. A framework analysis of the open source software development paradigm. In *ICIS*, Association for Information Systems (2000), 58–69.
27. Goeminne, M., and Mens, T. A comparison of identity merge algorithms for software repositories. *Science of Computer Programming 78*, 8 (2013), 971–986.
28. Gould, R. V. The origins of status hierarchies: A formal theory and empirical test. *Am. J. Sociol. 107*, 5 (2002), 1143–1178.
29. Gousios, G. The GHTorent dataset and tool suite. In *MSR*, IEEE (2013), 233–236.
30. Harrison, D. A., et al. Beyond relational demography: Time and the effects of surface-and deep-level diversity on work group cohesion. *Acad. Manag. J. 41*, 1 (1998), 96–107.

31. Horwitz, S. K., and Horwitz, I. B. The effects of team diversity on team outcomes: A meta-analytic review of team demography. *J. Manag. 33*, 6 (2007), 987–1015.
32. Howison, J., et al. Validity issues in the use of social network analysis with digital trace data. *JAIS 12* (2011).
33. Jackson, S. E., and Joshi, A. Diversity in social context: a multi-attribute, multilevel analysis of team diversity and sales performance. *J. Organ. Behav. 25*, 6 (2004), 675–702.
34. Kalliamvakou, E., et al. The promises and perils of mining GitHub. In *MSR*, ACM (2014).
35. Kouters, E., et al. Who's who in Gnome: Using LSA to merge software repository identities. In *ICSM* (2012), 592–595.
36. Lakhani, K. R., and Von Hippel, E. How open source software works: "free" user-to-user assistance. *Research Policy 32*, 6 (2003), 923–943.
37. Lee, M. J., et al. GitHub developers use rockstars to overcome overflow of news. In *CHI*, ACM (2013), 133–138.
38. Lipsey, M., et al. Translating the statistical representation of the effects of education interventions into more readily interpretable forms. Tech. Rep. NCSER 2013-3000, U.S. Dept of Education, Institute of Education Sciences, 2012.
39. Marlow, J., et al. Impression formation in online peer production: activity traces and personal profiles in GitHub. In *CSCW*, ACM (2013), 117–128.
40. Martins, L. L., et al. Virtual teams: What do we know and where do we go from here? *J. Manag. 30*, 6 (2004), 805–835.
41. McDonald, N., and Goggins, S. Performance and participation in open source software on GitHub. In *CHI*, ACM (2013), 139–144.
42. Michailidou, A., and Economides, A. Gender and diversity in collaborative virtual teams. *Computer Supported Collaborative Learning: Best practices and principles for instructors* (2007), 199–224.
43. Molleman, E., and Slomp, J. The impact of team and work characteristics on team functioning. *Hum. Factors Ergon. Manuf. 16*, 1 (2006), 1–15.
44. Nafus, D. 'Patches don't have gender': What is not open in open source software. *New Media & Society 14*, 4 (2012), 669–683.
45. Oreg, S., and Nov, O. Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Comput. Hum. Behav. 24*, 5 (2008), 2055–2073.
46. O'Reilly III, C. A., et al. Work group demography, social integration, and turnover. *Admin. Sci. Quart.* (1989), 21–37.
47. Preece, J., and Shneiderman, B. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Trans. Hum.-Comput. Interact. 1*, 1 (2009), 13–32.
48. Riordan, C. M., and Shore, L. M. Demographic diversity and employee attitudes: An empirical examination of relational demography within work units. *J. Appl. Psychol. 82*, 3 (1997), 342.
49. Robles, G., et al. FLOSS 2013: A survey dataset about free software contributors: challenges for curating, sharing, and combining. In *MSR*, ACM (2014), 396–399.
50. Robles, G., and Gonzalez-Barahona, J. M. Contributor turnover in libre software projects. In *Open Source Systems*. Springer, 2006, 273–286.
51. Rogelberg, S. G., and Rumery, S. M. Gender diversity, team decision quality, time on task, and interpersonal cohesion. *Small Group Research 27*, 1 (1996), 79–90.
52. Salancik, G. R., and Pfeffer, J. A social information processing approach to job attitudes and task design. *Admin. Sci. Quart. 23*, 2 (1978), 224–253.
53. Simmons, J., et al. False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci. 22*, 11 (2011).
54. Stahl, G. K., et al. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *J. Int. Bus. Stud. 41*, 4 (2010), 690–709.
55. Tajfel, H. Social psychology of intergroup relations. *Annu. Rev. Psychol. 33*, 1 (1982), 1–39.
56. Takhteyev, Y., and Hilts, A. Investigating the geography of open source software through GitHub, 2010.
57. Tao, Y., et al. How do software engineers understand code changes?: An exploratory study in industry. In *FSE*, ACM (2012), 51:1–51:11.
58. Turkle, S. *The Second Self: Computers and the Human Spirit*. MIT Press, 2005.
59. Vasilescu, B., Capiluppi, A., and Serebrenik, A. Gender, representation and online participation: A quantitative study. *Interacting with Computers* (2013), 1–24.
60. Vasilescu, B., et al. StackOverflow and GitHub: Associations between software development and crowdsourced knowledge. In *SocialCom*, IEEE (2013), 188–195.
61. Vasilescu, B., et al. How social Q&A sites are changing knowledge sharing in open source software communities. In *CSCW*, ACM (2014), 342–354.
62. Vasilescu, B., et al. On the variation and specialisation of workload–A case study of the Gnome ecosystem community. *Empir. Softw. Eng. 19*, 4 (2014), 955–1008.
63. Wang, L., et al. Searching for the goldilocks zone: trade-offs in managing online volunteer groups. In *CSCW*, ACM (2012), 989–998.
64. Watson, W. E., et al. Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Acad. Manag. J. 36*, 3 (1993), 590–602.
65. Weyuker, E. J., et al. Do too many cooks spoil the broth? using the number of developers to enhance defect prediction models. *Empir. Softw. Eng. 13*, 5 (2008), 539–559.
66. Williams, K. Y., and O'Reilly, C. A. Demography and diversity in organizations: A review of 40 years of research. *Res. Organ. Behav. 20* (1998), 77–140.
67. Ye, Y., and Kishida, K. Toward an understanding of the motivation of open source software developers. In *ICSE*, IEEE (2003), 419–429.
68. Zellmer-Bruhn, M. E., et al. When and how do differences matter? an exploration of perceived similarity in teams. *Organ. Behav. Hum. Dec. 107*, 1 (2008), 41–59.

View publication stats