2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)

Gender in Open Source Communities: Different Migration Patterns and Forms of Work

David Gray Widder
Institute for Software Research, Carnegie Mellon University
dwidder@cmu.edu

I. Introduction and Motivation

Open source software (OSS) runs much of the world's computing infrastructure [1]. To survive, OSS projects need a steady supply of (often volunteer) labor. Most OSS projects are open to newcomers in theory, but research has uncovered technical and social barriers that newcomers face when joining [2]. The process of evolving from a peripheral contributor to a core contributor is referred as "migration" [3]. Core contributors conduct most of the work and thus are important for the long term health of an OSS project [4], [5].

While the tech industry at large faces a gender diversity issue, with only around 25% of technical roles being held by women [6], the problem is more egregious in OSS, where recent research reports the number of women at 5% [7]. Ethnographic work suggests that because different kinds of work in OSS communities are valued differently, this contributes to this gender imbalance, because work which tends to be done by women is undervalued [8]. For example, contributing lines of code are often given highest value, above other important community maintenance work such as communication, relationship building, and mentorship: tasks nonetheless vital to the long term health of an OSS community. Consequently, I seek to investigate what kinds of work different genders conduct on their migration paths to leadership roles in OSS communities.

This abstract describes interviews with 15 OSS contributors who are part of their project's core team to investigate the **(1) different forms of work they participate in**, and **(2) the process by which they migrated from a peripheral to a core contributor**. The long term goal of this work is to interview non-men and contrast their experiences with those of men. Unfortunately, the pitifully low number of non-men in OSS communities means that they are "unicorns" and often vastly overburdened by requests to participate in research such as mine. To ensure that I do not waste the time of these valuable participants, I have begun my work by interviewing core men in OSS communities to collect comparative data while refining my study design. I describe the results of these interviews with respect to migration patterns and different forms of work, and outline plans to extend this to people of all genders.

978-1-5386-5541-2/18/$31.00 ©2018 IEEE

II. Approach

A. Participants

Instead of sampling across many different ecosystems which may have wildly different cultures [9], I elect to focus on the PyPi ecosystem for consistency.[1] I use the traditional measure of "core" contributors: those contributors who have contributed more than given threshold of code commits (5% in this case) to their project [10]. Using the GHTorrent database [11], I identified candidate participants, and filtered out those who did not have public email addresses or had asked not to be contacted. Participants who were the sole contributor on their project were removed. 15 participants agreed to an email request to participate, out of 124 emails sent. This response rate (12%) is unsurprising given that highly active contributors receive many requests for participation.

B. Procedure and Analysis

Interviews were primarily conducted over the phone, but four participants with timezone or language constraints requested to participate over email. With phone participants, I explained the study and sought permission to record their responses. Over 30 to 60 minutes, I administered an interview protocol covering the forms of work they do on the project, and how the forms of their contributions have changed since joining the project. Using an iterative process, I then coded sections of the transcribed recordings and email responses relevant to the two topics of the research. Results are accompanied by associated participants' numbers (eg, "(3, 4)").

III. Results and Contribution

A. Different Forms of Work

My "core contributor" participants, identified using only their code commits recognize that shortcomings of this traditional definition. For example, participant 8 said "I think it's tempting to look purely at the number of commits or number of lines of code that you added, and that's a good metric sometimes. But for us it's more been about who's pulling our weight in terms of what we actually wanted to achieve".

Code review, often carried out by reviewing and commenting on pull requests, is one of the most mentioned forms of work (1, 2, 4, 5, 6, 9, 11, 12). Also, frequently mentioned is work helping evaluate and resolve issues and bug reports (2, 5, 6, 8, 11, 12)), and deal with other forms of external support

[1] https://pypi.org

requests using things like email or Stack Overflow (5, 7, 11, 12, 13). Others discussed debugging, maintaining and refactoring existing code (2, 6, 11, 13).

Many participants spoke of the outreach work they do such as presenting talks at academic or practitioner conferences (1, 4, 5, 7, 8, 11), teaching how to use the software at local universities (1, 11), and writing blog posts and promoting on social media (4, 5), among other activities (3). Similarly, others discussed creating documentation and tutorials (4, 8, 11, 12), making examples (8), writing academic papers about the project (7, 8). Many participants explicitly discussed recruiting, training, mentoring and encouraging new contributors (5, 6, 7, 9, 11, 12, 13), including new students (7, 9, 11, 13).

Participants also discussed how they plan the future of and make decisions on the project as part of their work. A few participants hold elected community management roles (1), or sit on a projects technical board (4). Others communicate with or manage other contributors, making decisions and coordinating work (2, 5, 12), using slack and other chat platforms (2, 5, 11), email lists (4, 8), in person (1, 2, 11), or on the phone (1).

B. Migration Patterns

Seven participants quickly became "core" on the project because they were hired to work on these projects (7, 9, 11), or because the project became a business need in their current work (1, 2, 10, 12). Six participants contribute as part of academic work, for example as part of their PhD (3, 8) or post doctoral research (9), as paid research software engineers (7, 11), or as part of a undergraduate robotics club (13). Three participants started the project I asked them about, thus becoming one of the primary contributors by default (5, 8, 12), some of which remain highly active with others now playing a supporting role. One does not identify as a core contributor (6), despite my sampling technique identifying him as so.

Some participants are no longer active members, whether because they have been reassigned at work (2), or have taken other jobs, or they consider the project to be stable and not needing many further contributions (12), or have finished their PhD and thus the project no longer relates to their work (3), and another will likely hand over the lead role on the project if he leaves academia after his PhD (8). One no longer writes new code, because his work has changed and because he considers the project stable, but continues to put in 20 minutes of work a week to support the projects community (5).

Some contributors appear to take on more of a community oriented role as their tenure with the project lengthens, encouraging new contributors, stepping back from writing new code and instead evaluating submitted pull requests (5), or by overseeing other paid contributors at the same workplace (2), but one describes the opposite: retreating from a community governance role to communicating with a select group of people and writing code (1).

IV. Limitations and Future Work

Because I only count *code* commits to determine core status as is standard in literature, I may be selecting those participants who prioritize code contributions, thus biasing my sample. Given what I have learned about the different forms of work in OSS, my next round of interviews will use a more holistic measure of core status, including things such as code review, community document maintenance, and support actions, much of which is also visible in trace data. However, other activities including teaching, outreach and private email communications are less visible in trace data, so I will need to design methods to capture this work.

Very few people I spoke with appeared to become "core" through a gradual process. This is perhaps because participants seemed to find it difficult to remember a rich chronology of their history of involvement on the projects I asked them about, even when prompted. In the future, I may consider asking them to do more involved walkthroughs to help refresh their memory, with reference to artifacts or visualizations on GitHub as a memory aid, or asking people who have more recently become core contributors.

Clearly, the biggest piece left to future work is to interview and contrast these men's experiences with those of other genders. In the next iteration of the study, I will integrate the lessons learned above into a refined protocol, recruit women and non-binary participants, and administer this updated protocol. I will then contrast the experiences of men with those of other genders. In the longer term, I hope to use this work to build best practices and tooling to help OSS communities be more inclusive, and explore whether automated mentorship can help interested non-male contributors migrate to core roles.

References

[1] N. Eghbal, *Roads and bridges: The unseen labor behind our digital infrastructure*. Ford Foundation, 2016.

[2] I. Steinmacher, T. Conte, M. A. Gerosa, and D. Redmiles, "Social barriers faced by newcomers placing their first contribution in open source software projects," in *Proc. Conf. on Computer supported cooperative work (CSCW)*. ACM, 2015, pp. 1379–1392.

[3] C. Jergensen, A. Sarma, and P. Wagstrom, "The onion patch: migration in open source ecosystems," in *Proc. Conf. on Foundations of software engineering (FSE)*. ACM, 2011, pp. 70–80.

[4] M. Valiev, B. Vasilescu, and J. Herbsleb, "Ecosystem-level determinants of sustained activity in open-source projects: a case study of the pypi ecosystem," in *Proc. Conf. on Foundations of software engineering (FSE)*. ACM, 2018, pp. 644–655.

[5] C. Miller, D. G. Widder, C. Kästner, and B. Vasilescu, "Why do people give up flossing? a study of contributor disengagement in open source," in *Proc. Intl. Conf. on Open Source Systems (OSS)*. Springer, 2019, pp. 116–129.

[6] C. Ashcraft, B. McLain, and E. Eger, *Women in tech: The facts*. National Center for Women & Technology (NCWIT), 2016.

[7] R. S. Geiger, "Summary analysis of the 2017 github open source survey," *arXiv*, 2017.

[8] D. Nafus, "patches dont have gender: What is not open in open source software," *New Media & Society*, vol. 14, no. 4, pp. 669–683, 2012.

[9] C. Bogart, C. Kästner, J. Herbsleb, and F. Thung, "How to break an api: Cost negotiation and community values in three software ecosystems," in *Proc. Conf. on Foundations of software engineering (FSE)*. ACM, 2016, pp. 109–120.

[10] M. Joblin, S. Apel, C. Hunsen, and W. Mauerer, "Classifying developers into core and peripheral: An empirical study on count and network metrics," in *Proc. International Conf. on Software Engineering (ICSE)*. IEEE, 2017, pp. 164–174.

[11] G. Gousios and D. Spinellis, "Ghtorrent: Github's data from a firehose," in *Proc. Conf. Mining Software Repositories (MSR)*. IEEE, 2012, pp. 12–21.