

The State of Survival in OSS: The Impact of Diversity

Zixuan Feng
fengzi@oregonstate.edu
Oregon State University
Corvallis, Oregon, USA

ABSTRACT

Maintaining and retaining contributors is crucial for Open Source (OSS) projects. However, there is often a high turnover among contributors (in some projects as high as 80%). The survivability of contributors is influenced by various factors, including their demographics. Research on contributors' survivability must, therefore, consider diversity factors. This study longitudinally analyzed the impact of demographic attributes on survivability in the Flutter community through the lens of gender, region, and compensation. The preliminary analysis reveals that affiliated or Western contributors have a higher survival probability than volunteer or Non-Western contributors. However, no significant difference was found in the survival probability between men and women.

CCS CONCEPTS

• **Human-centered computing** → Open source software.

KEYWORDS

Open Source, Survivability, Disengagement

ACM Reference Format:

Zixuan Feng. 2023. The State of Survival in OSS: The Impact of Diversity. In *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE '23)*, December 3–9, 2023, San Francisco, CA, USA. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3611643.3617848>

1 INTRODUCTION

Open-source software (OSS) projects, a critical link in today's digital infrastructure, are volunteer-led, interest-driven, and comprised of contributors from diverse backgrounds [4, 6, 7]. Retaining and keeping contributors engaged is crucial in ensuring the sustainability of OSS projects [3, 7, 16, 27]. It is not surprising that the issue of retaining contributors in OSS has gathered significant attention, leading to extensive research focusing on contributors' survivability in OSS projects [3, 9, 15, 16, 23].

The survivability of OSS contributors can be impacted by their diverse backgrounds [3, 9, 15, 16, 23, 24]. Existing works have considered contributors' gender when investigating survivability in OSS projects [16, 24]. However, there is growing recognition that diversity encompasses more than just gender [7, 8]. Unfortunately,

there is a lack of research that investigates other aspects of demographics that may impact the contributors' survivability. This study investigates: **RQ: How do contributors' diverse background affect their survivability in OSS projects?**

To answer this research question, this study empirically analyzed the impact of demographic attributes pertaining to gender, region, and compensation on the survivability of contributors in the Flutter community. We chose Flutter because (1) it has a history of contributions since 2005, (2) it has a large contributor base, and (3) it has well-documented guidelines [1]. Analyzing the Flutter community allows this study to compare the current state of contributors' survivability and its historical trends.

2 BACKGROUND AND APPROACH

Contributors' survivability can be impacted by multiple factors [10, 12, 20, 26, 28]. Lin et al. [10] found that contributors who primarily modify code and contribute to the project earlier have a higher probability of survival. Regarding diversity, woman contributors are more likely to disengage than man contributors [16]. Studies on challenges faced by OSS contributors now consider diverse attributes such as gender, compensation, location, and seniority [8, 13, 15], no research has investigated how diversity affects the survivability of contributors beyond gender. To fill the gap, this study investigates 6,714 contributors' survivability through multiple dimensions of diversity.

Data: The dataset in this study includes 31 repositories in the Flutter community between the time frame of April 2005 to May 2023 and includes 182,584 commits, 596,072 issues and issue comments, and 329,490 pull requests and pull request comments [1].

Preprocessing and Filtering: To analyze the contribution history of each contributor, we identify if they contributed, regardless of the type of contribution (e.g., commits or pull requests or issues, or comments). We calculated the survival days by subtracting the date of the latest contribution from the date of the earliest contribution. As our analysis required GitHub profile information, we only included contributors who provided their full name and email address in their GitHub profile. By applying this criterion, our dataset comprised 16,294 contributors. We then excluded one-time contributors who were active for only one day, giving us a dataset of 8,365 contributors. It's worth noting that a one-time contributor might use multiple accounts for a single contribution per account.

Gender Inference: To predict the gender of contributors, we used "Namsor" API [2], a name prediction tool that estimates the gender of a person's name based on geographic information. Previous studies have reported the accuracy of "Namsor" is over 90% [19, 21]. However, not all contributors provide their region information in their GitHub profile. We first used "Namsor" to predict the origin of the name and then used the predicted region with their full name for gender prediction. To mitigate prediction errors, the

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 the author(s) 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.

ESEC/FSE '23, December 3–9, 2023, San Francisco, CA, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0327-0/23/12...\$15.00

<https://doi.org/10.1145/3611643.3617848>

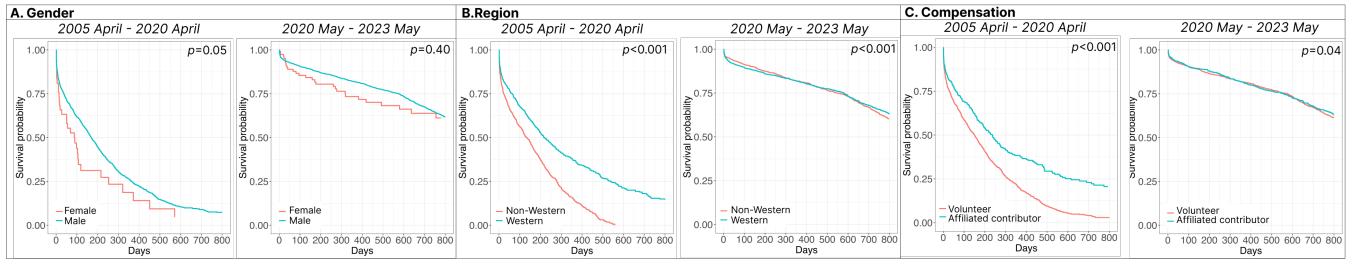


Figure 1: Kaplan-Meier curve disaggregated by diversity attributes (log-rank test p value shown in the graph).

final dataset included only names with a gender prediction probability over 90%, resulting in 6,534 male contributors and 180 female contributors.¹ It’s worth noting that contributors not categorized by “Namsor” due to our prediction threshold may be unrepresented in this study.

Region Inference: “Namsor” API analyzes a user’s full name to predict their country of origin. This feature provides information about the region of origin and the top 10 countries of origin [2]. We minimized errors by adopting and expanding a classification from a previous study [7], grouping contributors from 2,918 Western and 3,796 Non-Western countries. This classification considered cultural and linguistic differences, but a contributor’s name origin doesn’t always indicate their current residence [3, 7].

Compensation Inference: We used contributors’ email domains to identify the corresponding commercial companies of the developers manually [28]. Contributors are classified as volunteers if their email domain is not a company name, such as @gmail.com. Otherwise, the contributor is considered an affiliated contributor. Our dataset includes 1,712 contributors affiliated contributors and 5,002 volunteer contributors. The affiliated contributors come from various domains, including google.com, microsoft.com, etc.

Survival analysis: Kaplan-Meier curves enable the comparison of survival probability among groups of individuals by computing the survival probability of the occurrence of an event at a certain time point [18]. We used Kaplan-Meier curves to display survival probabilities for diversity attribute groups and applied the log-rank test to assess significant survival differences over time. We used a 1-year inactivity threshold to determine if a contributor left, allowing for possible short breaks [5].

The extensive history of Flutter allowed us to analyze the longevity of its contributors’ engagement. Our preliminary analysis shows an activity increase from 2020-2023 with 4,422 active contributors, up from 2,292 before 2020, possibly due to the COVID-19 pandemic’s impact, as seen in other projects [11, 22]. Following the increase in activity, we divided the dataset into two subsets: May 2020 to May 2023 for recent contributor survivability and April 2005 to April 2020 for historical survivability trends. We also excluded contributors outside the 5th to 95th percentiles to minimize outlier influence [16].

3 PRELIMINARY RESULTS

Figure 1 displays the Kaplan-Meier estimates for each diversity attribute. For each attribute, we did two analyses: (1) *historical data*

¹Namsor predicts gender based on name/region (male/female); these may not reflect an individual’s gender identity.

from April 2005 to April 2020, and (2) *recent data* from May 2020 to May 2023. The survival curves display the survival rates among groups (y-axis) over time, with the x-axis representing survival days.

How does gender affect contributors’ survivability? As shown in Figure 1A, analysis of *historical data* shows that male contributors had a higher probability of longer survival (log-rank test, $p = 0.05$). This finding is in line with the previous study [16]. However, we see no statistically significant differences between the two groups (log-rank test, $p = 0.40$) in the *recent data*. This could be attributed to a recent study that indicates an ongoing improvement in gender diversity [7].

How does region affect contributors’ survivability? Regional disparities, including cultural and language differences, can cause challenges for OSS contributors [3, 14]. Figure 1B reveals that contributors from Non-Western countries tend to disengage significantly earlier than contributors from Western countries in both time series (log-rank test: *historical data*: $p < 0.001$; *recent data*: $p < 0.001$). Notably, the survival curve gaps between Western and Non-Western contributors narrowed during the analysis for the *recent data*. This implies that barriers arising from regional disparity are improving, which is in line with a recent case study in Apache projects; Feng et al. [7] found a decline in region-related challenges.

How does compensation affect contributors’ survivability? Promoting voluntary participation is crucial for OSS projects, as volunteers bring diverse perspectives, skills, and expertise [17, 25]. Figure 1C reveals that volunteer contributors are generally more likely to disengage than affiliated contributors (log-rank test: *historical data*: $p < 0.001$; *recent data* $p = 0.04$). The *recent data* analysis showing narrowing gaps between the survival curves of affiliated contributors and volunteers suggests a potential improvement.

4 CONCLUSION AND FUTURE WORK

This preliminary study on how different diversity attributes affect survivability shows: (1) Improved survivability for women, though a gender gap remains, and (2) Western or affiliated contributors have a higher survival rate. These changes may stem from OSS popularity or interventions by OSS projects. As a next step, we will investigate the underlying reason for these changes.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No.: 1901031, Google-Award for Inclusion Research Program, and Namsor. I am grateful to Dr. Anita Sarma for her insightful guidance throughout this work.

REFERENCES

- [1] 2023. <https://github.com/flutter>
- [2] 2023. <https://namsor.app/about-us>
- [3] Sogol Balali, Igor Steinmacher, Umayal Annamalai, Anita Sarma, and Marco Aurelio Gerosa. 2018. Newcomers' barriers... is that all? an analysis of mentors' and newcomers' barriers in OSS projects. *Computer Supported Cooperative Work (CSCW) 27* (2018), 679–714.
- [4] Amiangshu Bosu and Kazi Zakia Sultana. 2019. Diversity and inclusion in open source software (OSS) projects: Where do we stand?. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE, 1–11.
- [5] Fabio Calefato, Marco Aurelio Gerosa, Giuseppe Iaffaldano, Filippo Lanubile, and Igor Steinmacher. 2022. Will you come back to contribute? Investigating the inactivity of OSS core developers in GitHub. *Empirical Software Engineering* 27, 3 (2022), 76.
- [6] InduShobha Chengalur-Smith, Saggi Nevo, and Pindaro Demertzoglou. 2010. An empirical analysis of the business value of open source infrastructure technologies. *Journal of the Association for Information Systems* 11, 11 (2010), 3.
- [7] Zixuan Feng, Mariam Guizani, Marco A Gerosa, and Anita Sarma. 2023. The State of Diversity and Inclusion in Apache: A Pulse Check. *arXiv preprint arXiv:2303.16344* (2023).
- [8] Mariam Guizani, Bianca Trinkenreich, Aileen Abril Castro-Guzman, Igor Steinmacher, Marco Gerosa, and Anita Sarma. 2022. Perceptions of the State of D&I and D&I initiative in the ASF. In *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society*. 130–142.
- [9] Renee Li, Pavithra Pandurangan, Hana Frluckaj, and Laura Dabbish. 2021. Code of Conduct Conversations in Open Source Software Projects on Github. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.
- [10] Bin Lin, Gregorio Robles, and Alexander Serebrenik. 2017. Developer turnover in global, industrial open source projects: Insights from applying survival analysis. In *2017 IEEE 12th International Conference on Global Software Engineering (ICGSE)*. IEEE, 66–75.
- [11] Onkar S Malgonde, Terence JV Saldanha, and Sunil Mithas. 2023. RESILIENCE IN THE OPEN SOURCE SOFTWARE COMMUNITY: HOW PANDEMIC AND UNEMPLOYMENT SHOCKS INFLUENCE CONTRIBUTIONS TO OTHERS' AND ONE'S OWN PROJECTS. *MIS Quarterly* 47, 1 (2023).
- [12] Courtney Miller, David Gray Widder, Christian Kästner, and Bogdan Vasilescu. 2019. Why do people give up flossing? a study of contributor disengagement in open source. In *Open Source Systems: 15th IFIP WG 2.13 International Conference, OSS 2019, Montreal, QC, Canada, May 26–27, 2019, Proceedings 15*. Springer, 116–129.
- [13] Marco Ortu, Giuseppe Destefanis, Steve Counsell, Stephen Swift, Roberto Tonelli, and Michele Marchesi. 2017. How diverse is your team? Investigating gender and nationality diversity in GitHub teams. *Journal of Software Engineering Research and Development* 5, 1 (2017), 1–18.
- [14] Gede Artha Azriadi Prana, Denae Ford, Ayushi Rastogi, David Lo, Rahul Purandare, and Nachiappan Nagappan. 2020. Including Everyone, Everywhere: Understanding Opportunities and Challenges of Geographic Gender-Inclusion in OSS. *ACM Transactions on Software Engineering and Methodology (TOSEM)* (2020).
- [15] Gede Artha Azriadi Prana, Denae Ford, Ayushi Rastogi, David Lo, Rahul Purandare, and Nachiappan Nagappan. 2021. Including everyone, everywhere: Understanding opportunities and challenges of geographic gender-inclusion in oss. *IEEE Transactions on Software Engineering* (2021).
- [16] Huilian Sophie Qiu, Alexander Nolte, Anita Brown, Alexander Serebrenik, and Bogdan Vasilescu. 2019. Going farther together: The impact of social capital on sustained participation in open source. In *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. IEEE, 688–699.
- [17] Mehvish Rashid, Paul M Clarke, and Rory V O'Connor. 2019. A systematic examination of knowledge loss in open source software projects. *International Journal of Information Management* 46 (2019), 104–123.
- [18] Jason T Rich, J Gail Neely, Randal C Paniello, Courtney CJ Voelker, Brian Nussenbaum, and Eric W Wang. 2010. A practical guide to understanding Kaplan-Meier curves. *Otolaryngology—Head and Neck Surgery* 143, 3 (2010), 331–336.
- [19] Lucía Santamaria and Helena Mihaljević. 2018. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science* 4 (2018), e156.
- [20] Andreas Schilling. 2014. What do we know about FLOSS developers' attraction, retention, and commitment? A literature review. In *2014 47th Hawaii International Conference on System Sciences*. IEEE, 4003–4012.
- [21] Paul Sebo. 2021. Performance of gender detection tools: a comparative study of name-to-gender inference services. *Journal of the Medical Library Association: JMLA* 109, 3 (2021), 414.
- [22] Darja Smitte, Anastasiia Tklich, Nils Brede Moe, Efi Papatheocharous, Eriks Klotins, and Marte Pettersen Buvik. 2022. Changes in perceived productivity of software engineers during COVID-19 pandemic: The voice of evidence. *Journal of Systems and Software* 186 (2022), 111197.
- [23] Bianca Trinkenreich, Igor Wiese, Anita Sarma, Marco Gerosa, and Igor Steinmacher. 2022. Women's participation in open source software: A survey of the literature. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 31, 4 (2022), 1–37.
- [24] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark GJ van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. 2015. Gender and tenure diversity in GitHub teams. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3789–3798.
- [25] Bo Xu and Donald R Jones. 2010. Volunteers' participation in open source software development: a study from the social-relational perspective. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 41, 3 (2010), 69–84.
- [26] Yuxia Zhang, Klaas-Jan Stol, Hui Liu, and Minghui Zhou. 2022. Corporate dominance in open source ecosystems: a case study of OpenStack. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 1048–1060.
- [27] Minghui Zhou and Audris Mockus. 2012. What make long term contributors: Willingness and opportunity in OSS community. In *2012 34th International Conference on Software Engineering (ICSE)*. IEEE, 518–528.
- [28] Minghui Zhou, Audris Mockus, Xiujuan Ma, Lu Zhang, and Hong Mei. 2016. Inflow and retention in oss communities with commercial involvement: A case study of three hybrid projects. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 25, 2 (2016), 1–29.

Received 2023-06-05; accepted 2023-08-11