Contribution Patterns in Open Source Software for Social Good: Dynamics, Individuals, and Impact

ZIHAN FANG, Vanderbilt University, USA
YUEKE ZHANG, Vanderbilt University, USA
THOMAS ZIMMERMANN, University of California, Irvine, USA
DENAE FORD, Microsoft Research, USA
YU HUANG, Vanderbilt University, USA

Open Source Software for Social Good (OSS4SG), a specialized segment within the Open Source Software (OSS) domain, is gaining increasing recognition for its focus on addressing societal challenges and delivering positive social impact. Learning about how contributors engage with OSS4SG is crucial to its sustainability, as the long-term success of these projects relies heavily on active and ongoing contributor participation. However, no study has yet examined the dynamics of contributors within OSS4SG. To fill this gap, we analyzed over 2.2 million commits made by 5,860 contributors to both OSS4SG and general OSS projects on GitHub, identifying contribution patterns and factors influencing sustained contribution to OSS4SG. We found that although OSS4SG contributors tend to show lower overall contribution intensity and shorter active lifespans, their activity during engaged periods is relatively more regular compared to contributions to general OSS. In addition, contributors from developing regions (e.g., Africa) or women are more likely to start with and continue contributing to OSS4SG, despite their overall contribution levels being lower than those of others. Based on these insights, we propose targeted strategies to increase contributions to OSS4SG projects to maximize their social impact to benefit society and harness their potential to foster broader participation in open source, ultimately enhancing the sustainability of the whole community.

CCS Concepts: • **Human-centered computing**; • **Social and professional topics** → **User characteristics**;

Additional Key Words and Phrases: Open Source Software, Open Source Software for Social Good, Contribution Patterns

ACM Reference Format:
Zihan Fang, Yueke Zhang, Thomas Zimmermann, Denae Ford, and Yu Huang. 2026. Contribution Patterns in Open Source Software for Social Good: Dynamics, Individuals, and Impact. *Proc. ACM Hum.-Comput. Interact.* 10, 2, Article CSCW010 (April 2026), 26 pages. https://doi.org/10.1145/3788046

1 Introduction

Open Source Software for Social Good (OSS4SG), a specialized category of Open Source Software (OSS), consists of projects aimed specifically at addressing societal challenges and delivering positive social impact [24]. OSS4SG is gaining increasing recognition for its emphasis on solving social issues through collaborative software development. GitHub,[1] a leading platform for open-source collaboration, has significantly contributed to the expansion of OSS by attracting a large community

[1]https://github.com/

Authors' Contact Information: Zihan Fang, Vanderbilt University, Nashville, Tennessee, USA, zihan.fang@vanderbilt.edu; Yueke Zhang, Vanderbilt University, Nashville, Tennessee, USA, yueke.zhang@vanderbilt.edu; Thomas Zimmermann, University of California, Irvine, Irvine, California, USA, tzimmer@uci.edu; Denae Ford, Microsoft Research, Redmond, Washington, USA, denae@microsoft.com; Yu Huang, Vanderbilt University, Nashville, Tennessee, USA, yu.huang@vanderbilt.edu.

ACM 2573-0142/2026/4-ARTCSCW010
https://doi.org/10.1145/3788046

of contributors who engage in activities such as coding, issue reporting, and improving software quality [45]. This has also supported the rise of OSS4SG. While GitHub allows projects to self-identify through user-defined topics (e.g., #social-good), we identify OSS4SG by its alignment with the 17 Sustainable Development Goals (SDGs).[2] The SDGs serve as a comprehensive global framework aimed at addressing pressing social, economic, and environmental challenges, promoting peace, prosperity, and sustainability [6]. This aligns more closely with our research focus on sustainability and provides a consistent, externally grounded definition, as established in a prior study [15] and on the official Digital Public Goods Alliance website.[3] For example, the Somleng project is classified as OSS4SG because it aims to save lives by providing communication access to some of the world's most remote and vulnerable communities, striving to make communication accessible to everyone,[4] aligning with the *good health and well-being* and *reduced inequalities* goals of the SDGs.

Different from general OSS, previous research has shown that contributors to OSS4SG are driven by the desire to create lasting social change, prioritizing societal impact over personal gains such as skill development or career advancement [24]. They are also more selective, choosing projects based on societal relevance, intended beneficiaries, and broader social impact [24], whereas general OSS contributors may focus more on technical interests or personal preferences [19]. Thus, OSS4SG projects may struggle to attract contributors' attention at the same level as general OSS projects, since their focus on social impact can sometimes overshadow technical aspects that are often more appealing to the broader OSS community. Contributors to OSS4SG projects also face challenges such as limited visibility for OSS4SG and concerns about its long-term sustainability [24]. Both academia and industry have undertaken various initiatives to enhance their visibility. For instance, researchers developed a maintainer dashboard that provides strategies for attracting and retaining open-source contributors, which has been validated as effective in the OSS4SG project [22]. Another study introduced a recommendation system that suggests OSS4SG projects to contributors based on their activity and project-related data [16]. In an educational context, a study implemented a lightweight intervention that significantly increased the likelihood of student developers participating in OSS4SG [15]. Additionally, platforms like GitHub have actively promoted socially driven projects, encouraging community participation.

While existing studies have primarily focused on increasing the visibility of OSS4SG projects and attracting new contributors, they have largely neglected the challenges related to contributor participation and long-term contribution. Understanding how contributors maintain their participation over time is essential, as the development and success of OSS projects depend heavily on sustained community participation [27, 42]. However, there is limited understanding of whether and how contribution patterns in OSS4SG differ from those in general OSS. Furthermore, the factors influencing participation and sustained contributions in OSS4SG have yet to be thoroughly examined. Exploring these aspects can help enhance participation strategies, improve project sustainability, and maximize the social impact of OSS4SG, which is crucial for attracting diverse contributors and ensuring the success of projects aimed at addressing societal and global challenges.

We began with a set of 210 seed projects, encompassing both OSS and OSS4SG on GitHub, and collected over 2.2 million historical commits from their 5,860 contributors. We then examined contributors' contribution patterns in OSS4SG compared to those in general OSS projects and identified the factors influencing their participation and long-term contributions. Our findings indicate that contributors tend to show lower contribution intensity and shorter active duration in

[2]https://sdgs.un.org/goals
[3]https://www.digitalpublicgoods.net/
[4]https://github.com/somleng/somleng

OSS4SG compared to general OSS. However, their OSS4SG contributions are relatively more regular over time. While overall contribution levels to OSS4SG projects remain low, contributors who begin their open-source journey with OSS4SG tend to maintain a relatively high level of participation with social good projects throughout their contributing lifespan. In addition, contributors from developing regions (e.g., Africa) and women exhibit greater participation in OSS4SG compared to other demographic groups and are more likely to begin their open-source contributions with OSS4SG projects, although their total contributions remain lower than those of other contributors. We claim the following contributions:

- A comprehensive analysis of the differences in contributors' contribution patterns between OSS4SG and OSS.
- The first study to analyze the full historical commit activity of contributors who have engaged in OSS4SG.
- Insights for developing customized onboarding and retention strategies for OSS4SG.

By analyzing the contribution patterns of OSS4SG contributors, we propose targeted strategies to foster greater engagement and sustained involvement in OSS4SG, while also leveraging their potential to encourage broader participation in the open-source ecosystem. Our findings may also inform a wider range of technology-for-social-good communities (e.g., ICT4D and Data Science for Social Good), which, despite differing in scope and context, may face similar challenges related to visibility, inclusivity, and contributor retention. These efforts not only aim to improve the sustainability of social good initiatives but also to enhance the overall impact of technology-driven approaches to advance social development.

2 Related Work

2.1 Open Source Software For Social Good

Open Source for Social Good (OSS4SG) refers to open source software projects that are specifically designed to address societal issues by benefiting a community [24]. Despite their importance, these projects do not receive as much visibility and awareness as general OSS projects [24]. Various strategies have been employed to enhance their visibility within both industry and academia. For example, researchers have created a maintainer dashboard that provides strategies for attracting and retaining open-source contributors, such as highlighting project goals like social good to attract diverse contributors. These recommendations have proven effective in making OSS4SG projects appear welcoming, which also encourages maintainers to actively engage with new contributors [22]. Another study introduced a recommendation system that automatically suggests OSS4SG projects to potential contributors based on their activity and project-specific data. This system has proven effective, although there remains scope for further improvement in the quality of project recommendations [16]. In an educational context, researchers have attempted a lightweight intervention to encourage future developers to participate in OSS4SG, finding a significant improvement in the likelihood of student developers contributing to OSS4SG [15]. Furthermore, GitHub actively encourages developer participation in repositories focused on addressing social challenges. Their Social Impact team empowers current and future developers, as well as nonprofits and non-governmental organizations to make lasting, positive contributions to society by leveraging GitHub's products, brand, resources, and workforce.[5] Despite efforts to raise awareness among developers about the OSS4SG community and encourage their participation, the differences in contribution patterns between OSS4SG and general OSS projects remain unclear, which could also impact the sustainability and success of these projects. To address this, we performed a comparative

[5]https://github.com/social-impact

analysis of contributors' behavior in OSS4SG and general OSS projects to uncover the distinct characteristics and trends specific to OSS4SG.

2.2 Contributors in Open Source Software

2.2.1 Contributor Dynamics in Open Source Software. In OSS development, contributors play a critical role in molding projects and fostering collaborative innovation [55]. These contributors undertake a variety of tasks, including the creation, modification, and optimization of software to enhance its functionality [1, 43]. This spans both non-technical (e.g., feedback and mentorship) and technical (e.g., coding contributions) aspects [57]. The progress and sustainability of OSS projects heavily rely on the contributions and active participation of community members [27, 42]. The success of OSS is heavily dependent on consistent contributions from developers, as these regular inputs and active community participation are linked to enhanced software quality. Conversely, infrequent participation can negatively impact the quality of the software [7]. However, sustaining participation presents substantial challenges, including high turnover rates and burnout, which may threaten the viability of these projects [20, 28]. Thus, the dynamics of contributor participation and its impact on project development have been the subject of prior research. For instance, since many developers leave after making a few contributions [26], studies have identified factors that lead to dropout among contributors in open source, including the popularity of the projects they work on, their transitional experiences, the timing of their contributions, and the level of their participation [40].

On the other hand, investigations into what drives contributors to stay in OSS projects have uncovered both internal (e.g., personal satisfaction, intellectual stimulation) and external (e.g.,career advancement) factors [12, 19, 39, 61]. Moreover, active participation in community discussions and issue tracking can enhance participation and reduce turnover rates [26], while effective communication with maintainers and experienced contributors is crucial for the successful integration of newcomers [63]. Additionally, technical factors such as the codebase complexity also significantly influence contributor participation [3, 11]. Despite these technical factors, social interactions within the community, such as mentorship and collaboration, are also emphasized as crucial for attracting and retaining contributors [2, 64]. Therefore, to maintain contributors' participation, several effective strategies have been proposed, including recognizing contributors' efforts, establishing clear pathways for contributions, and fostering a welcoming community environment [21, 54, 64]. However, it remains unclear whether the social value orientation of OSS4SG can alleviate common problems such as contributor turnover and burnout, thus sustaining long-term participation. This study aims to investigate this potential.

2.2.2 Diversity of Contributors in Open Source Software. As OSS projects evolve, they increasingly attract contributors from diverse backgrounds [10]. Previous studies have explored the challenges and opportunities that diversity presents for OSS onboarding and contributions, focusing on factors such as region and gender differences [13, 46, 56]. For example, although women are significantly underrepresented in OSS communities, those who do participate demonstrate comparable productivity to men [18], follow similar development trajectories, and often remain more engaged in projects than men [13]. However, gender-related bias persists, as previous research indicates that pull requests submitted by women on GitHub are less likely to be accepted than those submitted by men [53]. Regarding geographical influences, research has shown that while most motivations and barriers to contributions are consistent across regions [46], geographical biases also significantly affect the acceptance of contributions, such as pull requests [49, 50]. These insights, however, pertain to the general OSS community. Considering the distinct aim of OSS4SG in addressing social issues, it remains to be determined whether factors like region or gender affect contributors' onboarding, participation, and sustained contribution to these projects.

2.3 Commit Contribution Patterns in Open Source Software

Commits represent a developer's contribution to a project's codebase [30]. Understanding the commit contribution patterns of contributors is crucial for discerning the evolutionary trends in OSS projects, as well as the behaviors and collaborative patterns of developers [33]. These elements are fundamental in influencing the sustainability of OSS projects [62]. Mockus et al. investigated commit activities in OSS projects such as Apache and Mozilla and found that while a small core group of developers contributed the majority of the work, a larger group of occasional contributors also played a significant role in project development [41]. Another study examined the initial commit behaviors of developers in OSS projects, modeling and predicting their long-term participation. It found that willingness and environment are critical in determining whether a newcomer will remain as a long-term contributor [64]. A different study analyzed the distribution of commit intervals at both project and file levels across the projects' lifecycles and individual releases, revealing that these distributions typically follow power-law patterns, offering insights into the dynamics of OSS development processes [36]. With an understanding of contribution patterns, researchers explored methods that can foster long-term viability and growth within these projects to enhance the sustainability of OSS. For instance, Calefato et al. developed a novel approach to detect periods of inactivity among developers by examining their contribution patterns to projects [5]. Moreover, based on machine learning techniques, historical commit data from contributors has been used to predict which new contributors to OSS repositories are likely to become long-term participants [14]. Another study introduced a mathematical framework to analyze synchronous development in OSS projects by using the time-series data of commit and communication activities. This framework underscores the vital role of synchronized activities in distributed software development and offers insights into effective collaborative mechanisms that enhance OSS environments [60]. In this study, we analyzed the commit contribution patterns of contributors to OSS4SG projects, identifying distinct trends in their activity compared to general OSS.

3 Methodology

We collected complete commit histories of 5,860 contributors across 210 seed projects, resulting in a final dataset of their commit records, the associated GitHub project for each commit, and any available demographic data. We then analyzed their contribution patterns in OSS4SG projects compared to general OSS projects, as well as how participation in OSS4SG projects influences contributors' future activity. In addition, we investigated how contributor profiles (i.e., region and gender) affect their contributions to OSS4SG projects. We mainly aim to answer:

- RQ1: How do contribution patterns differ between OSS4SG and general OSS?
- RQ2: Does contributing to OSS4SG have a sustained influence on contributors' future participation in OSS4SG?
- RQ3: How do different contributor profiles influence their contribution patterns in OSS4SG compared to general OSS?

3.1 Data Collection

3.1.1 Seed Projects. As shown in Fig. 1, we began with 70 OSS4SG (N_{SG}=70) projects. Following established practices [15], we collected these OSS4SG projects from the Digital Public Goods

Alliance (DPGA),[6] which defines projects for social good based on the 17 SDGs.[7] This approach aligns closely with the definition of "social good" as recognized by computing practitioners, as highlighted in a recent study [15]. To confirm that the selected projects truly serve social good purposes, one of the authors manually verified all 70 OSS4SG projects. To construct a baseline of general OSS projects, we randomly sampled **two** sets of general OSS projects from GitHub, following the methodology employed in a prior study [24]. The first set consisted of 70 general OSS projects with codebase size similar to those of the OSS4SG projects (N_{CS} = 70), while the second set included 70 general OSS projects with a similar number of contributors as the OSS4SG projects (N_{CC} = 70). For example, if an OSS4SG project had ten contributors, we selected a general OSS project with the same number of contributors, consistently applying this approach to choose OSS projects of similar codebase size. Consequently, we assembled a dataset of 210 seed projects: (1) 70 OSS4SG projects with an average of 64 contributors and an average codebase size of 210,102 KB; (2) 70 general OSS projects matched by codebase size; and (3) 70 general OSS projects matched by contributor number. Kolmogorov-Smirnov tests verified that other project characteristics, such as programming language, project age, and number of active days, showed no significant differences between the matched pairs.

3.1.2 Contributors' Activity. In this study, we defined a contributor as someone who has made at least one recorded commit in a repository's history, since commits serve as the primary indicator of activity on GitHub [29]. On GitHub, commits represent changes to one or more files within a project, capturing various actions such as adding new code, fixing bugs, or updating documentation.[8] Each commit is assigned a unique identifier, known as a SHA or hash, which allows precise tracking of every modification.[9] Our analysis focuses primarily on commit contributions, as they are the most accurate indicator of a developer's contributions and reflect the progression of an OSS project [29]. As shown in Fig. 1, we collected the complete commit history of contributors associated with the seed projects on GitHub.

3.1.3 Data Sources. We used two sources for data collection: GitHub API and World of Code.

GitHub API. The GitHub API enables comprehensive access to GitHub's features, allowing researchers and developers to automate tasks such as repository management, issue tracking, and pull request handling.[10] The API supports both REST and GraphQL, providing flexibility in data retrieval and manipulation. Fig. 1 illustrates the role of the GitHub API in the data collection process. We primarily used the GitHub API to gather publicly available demographic information of contributors, including their regions and names listed on their GitHub profiles.

World of Code. World of Code (WoC) is a scalable and up-to-date infrastructure designed to support research using version control data from open-source projects, offering a comprehensive dataset that spans numerous repositories [34]. It provides broad access to its extensive dataset through a tool optimized for rapid query processing, improving data usability for a variety of research tasks [34]. We chose WoC due to its ability to efficiently collect and process large datasets at scale, as well as the limitations of the GitHub API in retrieving a user's complete commit history.[11] WoC's capabilities have been validated through its widespread application in diverse research [34, 35]. Fig. 1 shows how we used WoC in the data collection process. Specifically, we

[6] https://digitalpublicgoods.net/
[7] https://sdgs.un.org/goals
[8] https://github.com/git-guides/git-commit
[9] https://docs.github.com/en/pull-requests/committing-changes-to-your-project/creating-and-editing-commits/about-commits
[10] https://docs.github.com/en/rest?apiVersion=2022-11-28
[11] https://docs.github.com/en/rest/activity/events?apiVersion=2022-11-28

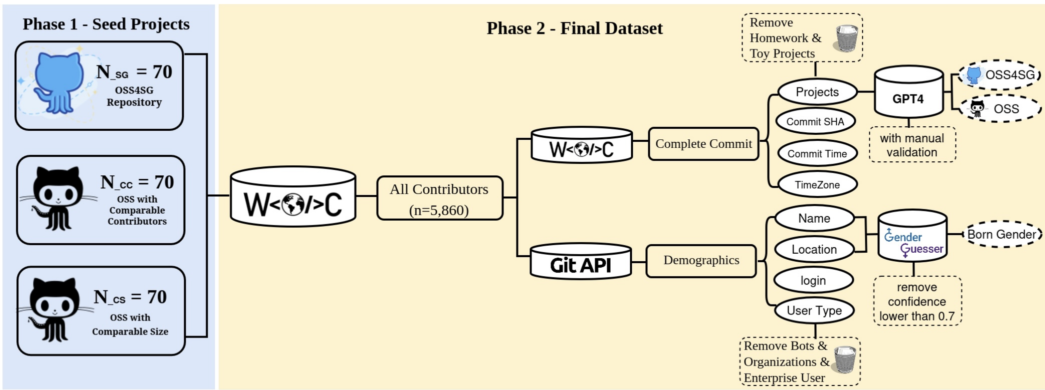

Fig. 1. Data Collection Process. Starting with 210 seed projects, the process results in a comprehensive dataset containing contributors' complete historical commits and demographic details.

used WoC to extract contributors associated with seed projects, along with their historical commit records. These commit records include details such as the commit SHA, timestamp, timezone, author, and the repository to which the commit is associated.

3.1.4 Data Collection. The data collection process, as outlined in Fig. 1, consists of two phases.

In **Phase 1**, we collected a total of 210 GitHub repositories as seed projects. These included 70 OSS4SG projects, 70 general OSS projects matched by contributor number, and 70 general OSS projects matched by codebase size.

In **Phase 2**, we began by using WoC to collect all contributors from the 210 seed projects, along with their complete historical commit records on GitHub. This dataset encompasses their commit activity not only in the seed projects but also across all other GitHub projects they have contributed to. Specifically, the collected data includes, for each commit: the commit SHA, timestamp, timezone, author, and the corresponding project associated with the commit. For each newly included project in this phase, we first employed GPT-4.0 for preliminary classification,[12] then conducted manual verification to assess whether the project should be classified as OSS4SG or general OSS. This process is further detailed in Sec. 3.2.2.

Moreover, since WoC does not provide demographic information about contributors, we supplemented the data by querying the GitHub API to retrieve additional details, including user type, region, and name publicly displayed on their GitHub profiles. Specifically, the process began by using the WoC user index (which consists of combined names and email addresses) to identify the public commits associated with each contributor through their email addresses. From these commits, we extracted the corresponding GitHub logins, which served as identifiers to access supplementary profile data via the GitHub API. To infer the likely gender of contributors, we used Namsor, a tool that predicts gender based on an individual's first name and the cultural origin of their last name.[13] Namsor has been recognized as one of the most accurate name-based gender inference methods in prior evaluations [51, 52]. Consequently, in this phase, we constructed two datasets: one comprising contributors' commit histories and another containing their profile information. These were subsequently merged to produce a final dataset with 11 attributes: WoC user index (i.e., author of this commit), commit SHA, associated project, project type (i.e., OSS4SG or general OSS), commit time, timezone, user's GitHub login, name, region, user type, and inferred gender. To

[12]https://openai.com/index/gpt-4/
[13]https://gender-guesser.com/

ensure privacy, the demographic information has been anonymized and stored securely. All data were collected as of May 2023.

3.2 Data Processing and Analysis

3.2.1 Data Processing and Validation. The final dataset consists of 11 attributes, which encompass the WoC user index, commit SHA, associated projects, and other relevant fields as detailed in the previous section. To ensure the reliability of our findings and address potential noise in the raw data, we implemented a comprehensive data cleaning process. We first excluded ghost users, whose GitHub login could not be retrieved via the GitHub API.[14] The first author manually confirmed that these contributors were indeed ghost users and no longer active on GitHub. These accounts represent 1.51% of the contributors collected in this phase. We also excluded contributors whose user types were Bots, Enterprises, or Organizations on GitHub, as they typically do not represent individual human contributors.[15]

To mitigate the influence of trivial projects such as toy examples or academic assignments, which generally do not reflect substantial software development efforts and could potentially distort the dataset, we implemented a two-stage filtering process. First, we identified and excluded homework projects by analyzing repository names and README file descriptions. Subsequently, adhering to established practices [25], we removed toy projects containing fewer than 100 commits. Additionally, to ensure temporal consistency across all records, we converted local timestamps to UTC using the time zone information provided in the WoC dataset. After implementing these filtering and standardization procedures, our final dataset consisted of 5,860 distinct contributors and 101,339 projects they have engaged in, totaling 2,276,926 valid commit records.

Among these contributors, 2,757 had valid region data, which we manually standardized as continents. The geographic distribution included 816 contributors from Asia, 1,082 from Europe, 659 from North America, 63 from Africa, 82 from South America, 54 from Oceania, and one from Antarctica. For contributors with publicly available name and region data, we inferred their likely gender using Namsor, which provides a confidence score indicating the reliability of its gender prediction. Consistent with the methodology employed in prior research [47], we excluded gender prediction with a confidence score below 0.7 to ensure reliability. After applying this threshold, we identified 211 women and 2,032 men.

3.2.2 OSS4SG Annotation. As mentioned in Sec. 3.1.4, in **Phase 2**, we collected the complete commit history of contributors, including the corresponding projects associated with each commit. Each newly identified project was classified as either OSS4SG or general OSS through a two-step process: automated GPT-4.0 classification followed by manual verification. Specifically, we used the GitHub API to retrieve project descriptions, which were then analyzed by GPT-4.0 to determine whether they aligned with any of the SDGs. Projects matching these criteria were labeled as OSS4SG. To ensure classification accuracy, all projects initially identified as OSS4SG underwent manual verification by two authors to confirm their social good orientation. For projects classified as general OSS by GPT-4.0, we performed quality control by having the first author manually examine a random sample of 2,000 projects. This manual examination revealed that the automated classification achieved an accuracy rate of 95.3%. In addition, two authors manually reviewed and labeled projects that lacked GitHub descriptions. Through this process, we identified 1,672 OSS4SG projects (52.2% associated with contributors from the 70 seeded OSS4SG projects and 47.8% from the matched OSS seed projects) and 99,667 general OSS projects within the final dataset.

[14]https://github.com/ghost
[15]https://docs.github.com/en/get-started/learning-about-github/types-of-github-accounts

3.2.3 Data Analysis. We detailed the methods used to analyze the data for each research question. The finalized dataset comprises multiple attributes, including commit SHA, associated project, and project type. Accordingly, a commit is classified as an OSS4SG contribution when it is sourced from a OSS4SG project.

RQ1: Of the 5,860 contributors in the final dataset, 75 exclusively contributed to OSS4SG projects, 974 participated in both general OSS and OSS4SG projects, while 4,811 were active only in general OSS projects throughout their GitHub activity. To investigate differences in contribution patterns between OSS4SG and general OSS, we focused our analysis on the 974 contributors who have engaged in both types of projects. For an initial assessment of how contribution intensity varies between OSS and OSS4SG, we compared each contributor's average number of projects and commits in OSS4SG to their corresponding averages in general OSS. We then investigated whether their contributions are temporally clustered or evenly distributed across their contribution lifespan. To quantify this behavior, we used the regularity metric (R), defined as:

$$R = \frac{\sigma - \mu}{\sigma + \mu}$$

where σ and μ represent the standard deviation and mean of inter-commit times, respectively. Inter-commit time refers to the number of days between a contributor's consecutive commits. Higher values of R ($R \to 1$) indicate irregular or sporadic contribution frequency, whereas lower values ($R \to -1$) suggest more regular and consistent contribution. Lastly, we measured contributors' active durations in both general OSS and OSS4SG projects by calculating the days between their first and last recorded commit for each project. To control for differences in project age, we normalized these durations by the project's total lifespan in days.

RQ2: To evaluate whether contributing to OSS4SG projects influences contributors' future activity on GitHub in general, we analyzed the complete cohort of 5,860 contributors. We began by examining whether the type of project (OSS4SG or general OSS) to which a contributor made their first contribution affected their subsequent contribution. We categorized contributors into two groups: OSS4SG starters (those whose first recorded contribution was to an OSS4SG project) and OSS starters (those whose first contribution was to a general OSS project). For both groups, we measured participation levels through average project participation and commit contribution. Moreover, we conducted a targeted analysis of 1,049 contributors with prior OSS4SG participation, which included 225 OSS4SG starters and 824 OSS starters. This analysis aimed to determine whether their initial contribution type influenced their future OSS4SG contributions. Specifically, we assessed their OSS4SG contribution by determining what percentage of their total projects and commits were dedicated to OSS4SG. Furthermore, we examined how these proportions changed over annual, monthly, and weekly periods following their first contribution.

We further investigated whether OSS4SG contribution promotes continued engagement within OSS4SG, independent of contributors' entry point. Using the same cohort of 1,049 contributors, we identified each person's first OSS4SG contribution year and grouped contributors accordingly. This approach allowed us to analyze how the proportion of OSS4SG commits per contributor evolved within each cohort.

RQ3: We investigated the influence of profiles (i.e., region and gender) on contribution to OSS4SG. Within our dataset, 2,757 contributors had publicly available and valid region information, and 2,243 had reliable gender data. Projects were associated with the regions and genders of their contributors, meaning that a single project could be linked to multiple regions and genders depending on the diversity of its contributor base. The same approach was applied to commits, with each commit uniquely associated with the region and gender of the contributor who made it.

First, we examined the regions and genders from which contributors were more likely to begin their open-source journey with OSS4SG. This was assessed by calculating the percentage of OSS4SG starters within each gender and region group. Next, to assess the potential impact of gender and region on contributions to OSS4SG, we analyzed the percentage of contributors from each gender and regional group who have participated in OSS4SG projects throughout their contribution history. Furthermore, we calculated the total proportion of OSS4SG projects and commits across gender and regional groups to assess overall contributions from a broader perspective. For more detailed insights, we analyzed the mean percentage of OSS4SG projects and commits per contributor within each group.

3.2.4 Statistical Analysis. We validated non-normally distributed data from independent samples using the Mann-Whitney U test, a non-parametric statistical method suitable for comparing group differences [38]. For non-normally distributed data within the same group, we employed the Wilcoxon signed-rank test [59]. To summarize central tendency, we reported both the mean to reflect average contribution levels and the median to provide a more robust measure of typical behavior in the presence of skewed data or outliers, along with the standard deviation to capture the degree of variability in contribution patterns. For instance, in **RQ1**, we employed the Wilcoxon signed-rank test to assess whether the same group of contributors exhibited significantly higher contribution intensity to general OSS compared to OSS4SG projects throughout their entire contribution history, given the non-normal distribution of the data. In **RQ2**, we used the Mann-Whitney U test to compare contribution differences between OSS starters and OSS4SG starters.

Moreover, to assess whether differences exist across multiple groups (e.g., by region), we conducted chi-square tests [17]. Upon finding a significant association, we performed post hoc analysis using pairwise proportion z-tests or Mann-Whitney U tests with Bonferroni correction to account for multiple hypothesis testing [4]. For example, in **RQ3**, we examined whether the proportion of contributors who had previously participated in OSS4SG projects varied significantly across regions using a chi-square test. We then performed pairwise proportion z-tests with Bonferroni correction to determine whether the proportion of OSS4SG contribution from one region was significantly higher than that of another.

Furthermore, for **RQ2**, we used logistic mixed-effects regression to examine whether contributors sustain their OSS4SG activity over time. We model the probability that a contribution is to an OSS4SG project as a function of the elapsed time (in years) since the contributor's first OSS4SG commit, using a generalized linear mixed-effects model with a logit link. The model includes random intercepts for contributors and projects to account for repeated observations and the cross-classified structure of the data. In this specification, the fixed-effect coefficient for elapsed time captures the change in the log-odds (and corresponding odds ratio) that a contribution is to OSS4SG for each additional year since the first OSS4SG commit. We estimate the model via maximum likelihood and summarize uncertainty using standard errors, Wald z-statistics, and p-values. As a robustness check, we also fit a Bayesian logistic mixed-effects model on a stratified subset of the data and report posterior odds ratios with 95% credible intervals.

4 Results

We begin by analyzing differences in contribution patterns between OSS4SG and general OSS projects among contributors who have participated in both. We then assess the long-term impact of contributing to OSS4SG on continued participation in open source, both broadly and within OSS4SG specifically. Finally, we examine how distinct contributor profiles correspond to different patterns of OSS4SG participation and contribution.

4.1 RQ1: How do contribution patterns differ between OSS4SG and general OSS?

We analyzed the commit histories of 974 contributors who have contributed to both general OSS and OSS4SG projects on GitHub to investigate differences in their contribution patterns. We aim to assess whether contributors exhibit similar levels of sustained contribution to OSS4SG as to general OSS.

4.1.1 General Contribution Patterns in OSS4SG and OSS. We first examined the overall participation and contribution levels of contributors to OSS4SG compared to general OSS projects. *We found that contributors generally exhibited lower participation and contribution intensity in OSS4SG compared to general OSS.* Specifically, on average, they participated in significantly more general OSS than OSS4SG projects throughout their contribution history ($p < 0.001$), with a mean of 59.0 general OSS projects (median = 11.0) compared to 2.6 OSS4SG projects (median = 1.0) per contributor. Consequently, across their entire contribution history, contributors made significantly more commits to general OSS projects than to OSS4SG projects ($p < 0.001$), with an average of 861.9 commits per person (median = 128.5) in general OSS compared to only 64.1 commits on average (median = 6.0) in OSS4SG. This pattern remained even after being normalized by the number of projects. When examining the average number of commits per project per contributor, we still found that they generally contribute more to general OSS projects, averaging 23.4 commits per project (median = 9.4), compared to 24.3 commits per OSS4SG project (median = 4.0). Although the mean was slightly higher for OSS4SG, the higher median and distribution of values indicate greater contributor activity in general OSS overall, as confirmed by a Wilcoxon signed-rank test ($p < 0.001$). These findings indicate that contributors who engage in both types of projects generally prioritize general OSS as their main focus while treating OSS4SG as a side interest.

Building on the observation that contributors demonstrate varying levels of contribution intensity between OSS4SG and general OSS projects, we next examined whether their contribution frequency varies across project types. *We found that contributors to OSS4SG projects tend to maintain more regular contribution frequency, whereas contributions to general OSS projects are more clustered in short bursts.* To characterize contribution frequency, we employed the regularity metric (R) as defined in Sec. 3.2, where higher values ($R \rightarrow 1$) indicate more irregular contributions and lower values ($R \rightarrow -1$) signify more regular contribution patterns. Our analysis reveals that among contributors who participated in both types of projects, the mean regularity for OSS4SG contributions was 0.44 (median = 0.47), while general OSS contributions showed a mean regularity of 0.57 (median = 0.60), as shown in Table 1. This difference was statistically significant, as confirmed by a Wilcoxon signed-rank test ($p < 0.001$). The more regular but less intensive contributions to OSS4SG projects may indicate that contributors may be driven by sustained, mission-oriented motivations, such as a desire for social impact or a sense of civic responsibility, which foster regular engagement over time.

Table 1. Comparison of per-contributor participation and commit contribution in general OSS vs. OSS4SG (N = 974). The differences are statistically significant, confirmed by the Wilcoxon signed-rank tests.

Metric	**OSS**			**OSS4SG**		
	Mean	*Median*	*SD*	*Mean*	*Median*	*SD*
Average Projects Participated	59.0	11.0	424.9	2.6	1.0	5.8
Average Commits Contributed	861.9	128.5	3452.6	64.1	6.0	239.0
Average Commits per Project	23.4	9.4	43.8	24.3	4.0	67.8
Commit Regularity (R)	0.57	0.60	0.21	0.44	0.47	0.22
Normalized Active Duration	0.06	0.01	0.12	0.05	0.00	0.09

Note: All differences are significant at $p < 0.001$.

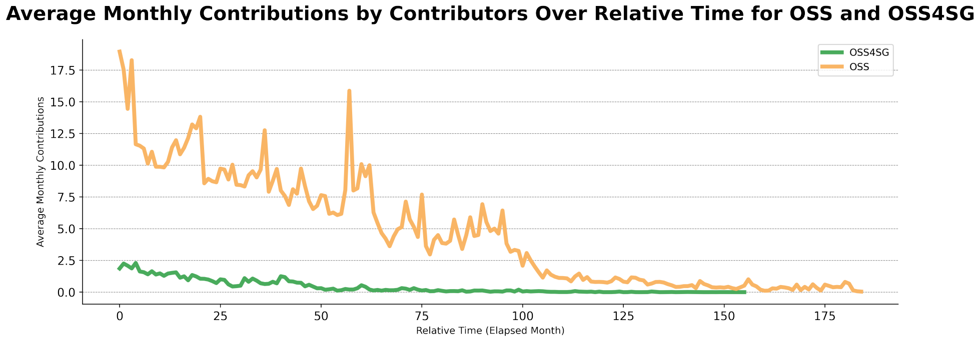

Fig. 2. The average number of OSS4SG commit contributions made by contributors monthly. The x-axis indicates elapsed months since their first open-source commit contribution (e.g., 25: 25th months after first commit contribution).

4.1.2 Long-term Contributions in OSS and OSS4SG. Building on our understanding of contributors' overall activity across their contribution lifespan, we next analyze how their engagement in OSS4SG evolves compared to general OSS. *We found a clear decline in contributions across both types of projects as time elapsed.* As shown in Fig. 2, contributors generally exhibit a decline in contributions over time to both OSS and OSS4SG projects, though the trajectories differ. They tend to contribute more actively to general OSS projects at the outset, with noticeable fluctuations marked by several peaks and dips, before gradually declining to minimal levels. In comparison, their contributions to OSS4SG decrease more steadily, reaching low levels more rapidly with less variability. This pattern persists consistently across multiple temporal scales (e.g., yearly and weekly). The more rapid decrease in OSS4SG contributions raises concerns about the long-term sustainability of these projects and highlights the need for targeted interventions to maintain continued contributions.

To better understand the decline in contributions, we examined the durations of contributor activity, comparing how long they remained active in general OSS compared to OSS4SG. *We found that contributors tend to have significantly shorter active durations in OSS4SG projects.* As detailed in Sec. 3.2.3, we defined a contributor's active duration as the number of days between their first and last recorded contributions, normalized by the corresponding project's lifespan (in days). On average, contributors maintained activity in general OSS projects for a normalized duration of 0.06 (median = 0.01), whereas when contributing to OSS4SG projects, they remained active for a shorter duration of 0.05 (median = 0.00), as presented in Table 1. A Wilcoxon signed-rank test confirmed this difference to be statistically significant ($p < 0.001$). These further suggest that OSS4SG may face greater sustainability challenges, potentially due to constraints such as limited funding, organizational support, or community infrastructure when compared to general OSS projects.

Contributors engaged in both OSS and OSS4SG projects show consistently lower levels of participation and contribution intensity in OSS4SG compared to general OSS throughout their contribution lifespan. While their OSS4SG activity tends to follow a more regular contribution frequency, it declines more rapidly and stabilizes at lower levels than their general OSS activity. Moreover, these contributors maintain significantly shorter active durations in OSS4SG relative to their engagement in OSS.

4.2 RQ2: Does contributing to OSS4SG have a sustained influence on contributors' future participation in OSS4SG?

We examined whether initiating open-source contributions through OSS4SG affects contributors' broader open-source contribution patterns, as well as their specific engagement with OSS4SG. Furthermore, we explored how initial contributions to OSS4SG influence long-term engagement, aiming to identify the factors that may promote sustained contributions to OSS4SG projects.

4.2.1 Impact of First Contribution Type on Subsequent Contributions. Among all 5,860 contributors collected, 225 started open-source contributions through OSS4SG (OSS4SG starters), while 5,635 began with general OSS (OSS starters). We first explored how general contribution patterns differ between these two groups. *We found that although OSS4SG starters participate in fewer projects and make fewer total commits overall, statistical tests indicate that their per-project contribution intensity is comparable to that of OSS starters.* On average, OSS starters contributed to 21.6 projects (median = 6.0) and made 414.7 commits on average (median = 35.0) per person across their entire contribution lifespan ($p < 0.001$), whereas OSS4SG starters participated in an average of 7.0 projects (median = 3.0) and made 165.2 commits on average (median = 18.0). The differences in both project participation and total commits between the two groups are statistically significant ($p < 0.001$). However, when examining the average number of commits made by each contributor to each project, we found no significant difference between the two groups ($p = 0.143$) and both OSS starters and OSS4SG starters demonstrate comparable contribution intensity for each project they have contributed to. Specifically, OSS4SG starters made an average of 20.1 commits per project (median = 6.0), while OSS starters averaged 15.3 commits per project (median = 4.9).

To further analyze contribution patterns within OSS4SG, we examined whether OSS4SG starters maintained a higher level of sustained contribution to OSS4SG projects compared to OSS starters. Given the variation in total commit numbers and project participation among contributors, we employed a relative metric: the percentage of a contributor's activity devoted to OSS4SG, calculated as the proportion of OSS4SG contributions relative to their total contributions. To account for the fact that many general OSS starters never contributed to OSS4SG projects, our analysis focused specifically on contributors with at least one OSS4SG contribution (n = 1,049), consisting of 225 OSS4SG starters and 824 general OSS starters, as shown in Table 2. *We found that OSS4SG starters demonstrate a stronger overall tendency to engage with OSS4SG than general OSS starters across their entire contribution lifespan.* On average, OSS4SG starters maintain 58.4% participation in OSS4SG projects (median = 50.0%) over their lifespan, compared to only 16.4% (median = 11.1%) for general OSS starters. This pattern is similarly reflected in their percentage of OSS4SG commit contributions, with OSS4SG starters contributing 64.2% (median = 73.2%) of their commits to OSS4SG projects, significantly higher than the 16.0% (median = 5.7%) for general OSS starters. The Mann-Whitney U test confirmed that these differences in both OSS4SG project participation and commit contributions between the two groups were statistically significant ($p < 0.001$).

To explore whether starting in OSS4SG alone explains this tendency, we classified contributors into four trajectories as shown in Table 3: Only-OSS4SG, who start and remain in OSS4SG; Start-SG→OSS, who start in OSS4SG and later expand to general OSS; Start-OSS→OSS4SG, who start in OSS and later expand to OSS4SG; and Only-OSS, who start and remain in general OSS. *We found that contributors who continue to invest substantially in OSS4SG are predominantly those who start in OSS4SG and later expand into broader OSS, not those who remain only in OSS4SG or arrive there after starting in general OSS.* Specifically, contributors who start in OSS4SG and later expand into OSS work on significantly more projects on average than Only-OSS4SG contributors (9.4 vs. 1.4, $p < 0.001$) and make substantially more commits overall (233.7 vs. 10.8, $p < 0.001$), while retaining sizable shares of their activity in OSS4SG (38.4% of projects and 46.7% of commits). They also exhibit

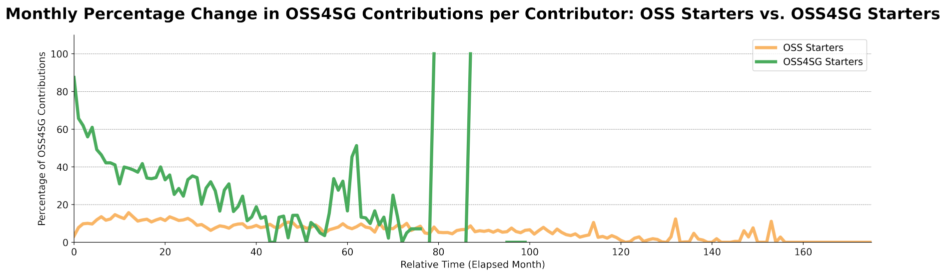

Fig. 3. Evolution of contributors' OSS4SG commit percentages over their contributing careers. The x-axis shows elapsed months since each contributor's first open-source contribution (e.g., 20 = the 20th month).

high per-project intensity (27.2 commits per project), comparable to those who start in OSS and later expand to OSS4SG (24.3 commits per project). By contrast, contributors who start in OSS and later expand to OSS4SG devote a smaller fraction of their activity to OSS4SG (18.6% of projects and 18.2% of commits), and contributors who only remain in OSS4SG show low overall activity despite their exclusive focus on social good projects. These suggest that OSS4SG sustainability depends on broader OSS trajectories in which social-good projects remain a substantial part of contributors' OSS participation.

Building on this, we next examine how contributions to OSS4SG evolve for both OSS and OSS4SG starters. *We found that OSS4SG starters' participation declines more rapidly, and they tend to leave earlier than OSS starters.* Specifically, OSS4SG starters begin by devoting a substantial share of their activity to OSS4SG projects, but this commitment decreases quickly and often ends prematurely. In contrast, general OSS starters maintain consistently low levels of OSS4SG contributions, which gradually decrease over time yet persist for a longer duration as shown in Fig. 3. A similar trend is observed in the percentage of OSS4SG projects involved. These findings suggest that while OSS4SG starters may have an initial willingness to engage in OSS4SG, they are not effectively retained within these projects. Similarly, OSS starters rarely shift a greater share of their future contributions toward OSS4SG.

Taken together, these results highlight a sustainability challenge for OSS4SG: the contributors who are both highly active and continue to engage meaningfully with OSS4SG are primarily those who start in OSS4SG and later expand into broader OSS, rather than those confined to OSS4SG or those who add OSS4SG after starting in general OSS. At the same time, OSS4SG starters as a group show faster declines in OSS4SG participation and leave earlier than OSS starters. However, given that OSS4SG starters account for only 21.5% of contributors who have contributed to OSS4SG, we further examined whether an initial contribution to OSS4SG, regardless of a contributor's starting point (i.e., OSS or OSS4SG), encourages continued engagement in social good projects.

4.2.2 Impact of OSS4SG Contribution on Subsequent Activity. We then examined *whether contributors are more likely to continue contributing to OSS4SG projects after their first OSS4SG contribution; however, we found no evidence to support this hypothesis.* Specifically, we modeled the probability that a contribution is to an OSS4SG project as a function of the years elapsed since the contributor's first OSS4SG commit, using a logistic mixed-effects regression with random effects for contributors and projects. Immediately after the first OSS4SG commit, the predicted probability that a subsequent contribution is to an OSS4SG project is about 35%, indicating a meaningful level of short-term engagement. However, this probability declines with each additional year ($z = -14.2$, $p < 0.001$). A Bayesian logistic mixed-effects model fitted on a stratified subset yields a similar pattern, with a

Table 2. Comparison of OSS and OSS4SG participation and commit contributions between OSS4SG starters and general OSS starters, normalized per contributor.

Metric	OSS4SG Starters (N = 225)			General OSS Starters (N = 5,635)		
	Mean	Median	SD	Mean	Median	SD
Average Projects Participated*	7.0	3.0	10.9	21.6	6.0	181.4
Average Commits Contributed*	165.2	18.0	413.7	414.7	35.0	4649.7
Average Commits Contributed Per Project	20.1	6.0	46.1	15.3	4.9	47.1
Metric	**OSS4SG Starters (N = 225)**			**General OSS Starters (N = 824)**		
	Mean (%)	Median (%)	SD (%)	Mean (%)	Median (%)	SD (%)
Percentage of OSS4SG Projects Participated*	58.4	50.0	33.5	16.4	11.1	15.3
Percentage of OSS4SG Commits Contributed*	64.2	73.2	35.6	16.0	5.7	21.4

* $p < 0.001$

Table 3. Comparison of participation metrics by contributor trajectory. Pairwise Mann–Whitney U tests with Benjamini–Hochberg correction show all group differences are significant ($p < .001$), except commits per project between the two expansion trajectories and projects participated between Start-SG→OSS and Only-OSS.

Metric	Only-OSS4SG	Start-SG → OSS	Start-OSS → OSS4SG	Only-OSS
Average Projects Participated	1.4	9.4	39.5	12.1
Average Commits Contributed	10.8	233.7	727.3	211.8
Average Commits Contributed Per Project	8.1	27.2	24.3	13.2
Percentage of OSS4SG Projects Participated	100.0	38.4	18.6	0.0
Percentage of OSS4SG Commits Contributed	100.0	46.7	18.2	0.0

Note. Only-OSS4SG: contributors who start and only remain in OSS4SG; Start-SG→OSS: start in OSS4SG and later expand to general OSS; Start-OSS→OSS4SG: start in general OSS and later expand to OSS4SG; Only-OSS: start and only remain in general OSS.

posterior mean odds ratio of 0.70 per additional year (95% CI: 0.65–0.76), implying roughly a 30% reduction in the odds of an OSS4SG contribution per year. The downward trend remains statistically significant after accounting for differences across contributors and projects via random effects, although there is substantial heterogeneity, suggesting that other factors likely drive variation in sustained engagement. These results indicate that OSS4SG does not become more central to contributors' portfolios; instead, its share of their contributions declines over time. External or individual factors influencing continued contribution therefore merit further investigation.

While OSS4SG starters demonstrate a stronger overall tendency to contribute to OSS4SG throughout their contribution lifespan and devote a larger share of their activity to social good projects, we find no evidence that the entry point (i.e., starting with OSS or OSS4SG) or contributing to OSS4SG necessarily leads to sustained engagement with these projects over time.

4.3 RQ3: How Do Different Contributor Profiles Influence Their Contribution Patterns in OSS4SG Compared to General OSS?

We examined how contributors' profiles influence their contribution patterns in OSS4SG to identify issues related to equity, accessibility, and representation in the sustainability of OSS4SG. However, due to limited demographic data, with valid region information for 2,757 contributors and gender information for 2,243, our analysis is restricted to these respective subsets.

4.3.1 Region Impact on First Contribution Type. Region bias is a known factor in OSS contributions, potentially affecting the acceptance of contributions on GitHub [48, 50]. However, it remains unclear whether the region is also a factor associated with contributors' participation in OSS4SG. We began by exploring the types of projects with which contributors from various regions are more likely to start. *We found that contributors from developing regions, such as Africa and South America, are more inclined to begin their open-source contribution with OSS4SG, compared to those from more developed regions like North America and Europe.* Specifically, Africa has the highest percentage of OSS4SG starters (9.5%), followed by South America (4.8%) and Asia (4.0%), as summarized in Table 4. A chi-square test confirmed the proportion of OSS4SG starters differs significantly across continents ($p = 0.024$). Subsequent pairwise z-tests with Bonferroni correction revealed that the higher proportion of OSS4SG starters in Africa is statistically significant compared to all other continents, except for South America. Additionally, the z-tests confirmed that South America's proportion is significantly higher than that of the remaining four continents. No statistically significant differences were observed among the other four continents. However, since the number of OSS4SG starters is small, these tests may be unstable; we therefore re-examined the association using a logistic regression with continent as a categorical predictor and a Bayesian logistic model with continent-level partial pooling. Both models replicated the pattern that African and South American contributors are more likely to start in OSS4SG than those from Europe and North America, with the Bayesian model stabilizing estimates for sparsely sampled regions. This suggests that regional factors, such as social needs or local challenges, may drive early participation in social good projects. Next, we investigated whether contributors from developing regions continue to maintain a higher level of participation in OSS4SG across their entire contribution lifespan.

Table 4. Percentage (%) distribution across continents. A statistically significant difference was observed between Africa and other continents across each metric. We **bolded** the highest value for each measurement.

Measurements	Africa	Asia	Europe	N. America	Oceania	S. America
Percentage of Contributors Starting with OSS4SG	**9.5**	4.0	2.5	3.6	0.0	4.8
Percentage of Contributors Who Have Contributed to OSS4SG	**47.6**	22.5	14.4	21.7	11.1	24.4
Overall Percentage of OSS4SG Projects Contributed to by Region	**7.4**	1.7	1.4	2.7	0.5	2.1
Average Percentage of OSS4SG Projects per Contributors	**10.5**	5.7	3.1	5.5	0.7	6.4
Overall Percentage of OSS4SG Commits Contributed to by Region	**14.5**	3.4	3.2	3.0	0.1	1.5
Average Percentage of OSS4SG Commits per Contributors	**13.7**	5.9	3.6	5.6	0.3	6.3

4.3.2 Region Impact on Contribution Patterns. When examining regional participation with OSS4SG, *we found that contributors from Africa exhibit a significantly higher propensity to participate in OSS4SG compared to those from other regions.* Specifically, Africa has the highest proportion of contributors with prior OSS4SG contributions (47.6%), followed by South America (24.4%) and Asia (22.5%). A chi-square test confirmed that a significant difference exists across the groups ($p < 0.001$). Paired z-tests further substantiate that the proportion of OSS4SG contributors in Africa is significantly higher than that of any other continent.

Moreover, at a broader scale, the overall proportions of OSS4SG projects and commits in each region further reinforce this observed pattern. Africa leads in the proportion of OSS4SG projects among all OSS projects contributed by the region, at 7.4%, followed by South America (2.7%) and Asia (2.1%). Likewise, the percentage of OSS4SG commit contribution is also highest in Africa (14.5%), ahead of Asia (3.4%) and Europe (3.2%). The statistical significance of Africa's elevated OSS4SG project participation and contribution rates is confirmed by both paired z-test and chi-square tests ($p < 0.001$).

To investigate this trend further, we conducted a contributor-level analysis examining the proportion of OSS4SG project participation and commit contributions per contributor, disaggregated by continent. From this finer-grained perspective, African contributors again show the highest likelihood of participation. On average, contributors from Africa participate in OSS4SG projects at a rate of 10.5%, followed by South America (6.4%) and Asia (5.7%). A similar trend is observed in commit contributions, where African contributors exhibit the highest average percentage of OSS4SG commits, accounting for 13.7% of their total commits followed by South American (6.3%) and Asian contributors (5.9%). The differences are statistically significant, as confirmed by pairwise Mann-Whitney U tests with Bonferroni correction for multiple comparisons ($p < 0.001$). Details of these findings are presented in Table 4. Thus, while the absolute number of contributions remains highest in more developed regions such as North America, contributors from developing regions are more likely to engage in OSS4SG and maintain a higher level of contribution throughout their contribution lifespan on GitHub.

4.3.3 Gender Impact on First Contribution Type. Gender is another common factor often considered when examining activity in OSS. Prior research has highlighted the presence of gender bias on GitHub, as well as gender-based disparities in contribution behaviors and outcomes within OSS communities [18, 44, 53]. Therefore, we investigated how gender relates to the types of projects contributors are more likely to join initially. *We found that women are significantly more likely to start their open-source contributions with OSS4SG.* Specifically, 6.2% of women made their first open-source contribution to OSS4SG projects, compared to 3.3% of men, with the difference being statistically significant (p =0.037). This may suggest a potential gender-based preference in initial project selection. We next examined whether this trend persists throughout their entire contribution life cycle.

Table 5. Percentage (%) distribution across genders. * indicates statistically significant differences between genders. We **bolded** the highest value for each measurement.

Measurements	**Men**	**Women**
Percentage of Contributors Starting with OSS4SG	3.3*	**6.2***
Percentage of Contributors Who Have Contributed to OSS4SG	18.3*	**28.9***
Overall Percentage of OSS4SG Projects Contributed to by Gender	1.7*	**2.9***
Average Percentage of OSS4SG Projects per Contributors	4.4*	**8.9***
Overall Percentage of OSS4SG Commits Contributed to by Gender	**3.8***	2.7*
Average Percentage of OSS4SG Commits per Contributors	4.7*	**9.7***

4.3.4 Gender Impact on Contribution Patterns. When examining how contributors of different genders participate in and contribute to OSS4SG throughout their contribution lifespan, *we found that women demonstrate a significantly stronger tendency to both participate in and contribute to OSS4SG.* Specifically, the percentage of women who have contributed to OSS4SG is significantly higher than that of men, as confirmed by a z-test (28.9% vs. 18.3%, p <0.001).

From a broader perspective, when analyzing the proportion of OSS4SG projects among all OSS projects contributed to by each gender, women demonstrate a significantly higher rate of participation than men (2.9% vs. 1.7%, $p < 0.001$). However, when examining contributions, the overall percentage of commits made to OSS4SG projects is significantly higher among men than among women (3.8% vs. 2.7%, $p < 0.001$). To gain a more granular understanding, we further examined the percentage of OSS4SG projects and commits per contributor, disaggregated by gender. On average, women exhibit greater engagement with OSS4SG projects. Specifically, the average percentage of OSS4SG projects per men is 4.4%, whereas it is 8.9% for women, a statistically significant difference confirmed by the Mann–Whitney U test ($p < 0.001$). Similarly, women contribute a significantly higher proportion of OSS4SG commits per person compared to men (9.7% vs. 4.7%, $p < 0.001$). Results are summarized in Table 5. Despite these relative differences, men still make more absolute contributions to OSS4SG projects. This disparity can be explained by two factors: men's generally higher levels of activity on GitHub and their larger overall presence on the platform. Taken together, these findings suggest that although women contribute less in absolute terms, they demonstrate a comparatively stronger inclination to engage with and contribute to OSS4SG projects.

> Contributors from developing regions, especially Africa, more often start and make contributions to OSS4SG, despite contributing less in absolute numbers than those from developed regions. Furthermore, women are more likely than men to begin their open-source journey with OSS4SG projects, and they tend to dedicate a higher proportion of their commits to these projects.

5 Threats to Validity

5.1 Internal

We began by collecting 70 OSS4SG projects from DPGA and randomly sampling 140 general OSS projects. Although we followed prior work by constructing comparable OSS baselines and verifying that observable characteristics were reasonably aligned, we acknowledge that more sophisticated matching (e.g., propensity-score matching on richer project-level covariates) would provide a more rigorous basis for comparison. Moreover, this seed sample may still represent only a subset of the broader OSS ecosystem. To mitigate this limitation, we used these 210 repositories solely as seed projects to identify contributors and then expanded our dataset to include all projects they contributed to, yielding 1,672 OSS4SG projects and 99,667 general OSS projects that better reflect the natural imbalance between OSS4SG and general OSS and provide broader coverage of the OSS ecosystem.

Moreover, given that our final dataset encompasses over 100,000 projects, including both OSS and OSS4SG projects, ensuring complete comparability across all projects (e.g., in terms of codebase size, domain, and number of contributors, etc) presents a significant challenge. This inherent variability may introduce concerns when comparing contributor behavior between OSS and OSS4SG projects. However, our goal is to capture contribution patterns within real-world open source ecosystems, where OSS and OSS4SG projects actually differ significantly in characteristics and scale. While this variation poses a limitation, it reflects the inherent diversity of the ecosystems we are studying. The broad scope of our dataset further mitigates the risk that our findings are influenced by a narrow or unrepresentative subset of projects, ensuring a more comprehensive and reliable understanding of contributor behavior. Nonetheless, we acknowledge that these differences could still influence their behavior. Furthermore, our longitudinal analyses are based on contributor histories observed within a finite time window, and OSS4SG projects may generally be younger than the comparison projects. As a result, estimates for contributors with longer contribution histories are supported by relatively few individuals and are best viewed as descriptive patterns rather than fully censoring-aware estimates (as shown in Sec. 4.2); more formal survival or event-history models would be a valuable extension. We therefore interpret our findings with this context in mind, recognizing the potential impact of ecosystem differences and censoring on the results.

In addition, our final dataset comprises 5,860 valid contributors. However, this sample may not fully capture the broader GitHub contributor population, as only approximately half of these contributors have publicly accessible profile information. As a result, our demographic findings should be interpreted as characterizing contributors with sufficiently complete public profiles and may underrepresent groups whose privacy preferences or profile-disclosure practices differ. This limitation may constrain the generalizability of our demographic-related analysis.

5.2 External

In **Phase 2** of data collection, we collected the complete historical commits of all contributors, including the projects associated with these commits. For the newly included projects, we classified them as either OSS4SG or general OSS projects using a hybrid approach that combined GPT-4.0-assisted classification with manual verification. However, we recognized the potential for misclassification by GPT-4.0, which may have led to some OSS4SG projects being incorrectly labeled as general OSS projects, and vice versa. To mitigate this risk, two authors manually reviewed all projects initially classified as OSS4SG to filter out the false positive classification. Furthermore, to assess the accuracy of GPT-4.0's classifications for general OSS projects, one author manually examined a random sample of 2,000 such repositories to confirm that they were not mistakenly labeled and did not meet the criteria for OSS4SG. This manual verification yielded an estimated classification accuracy of 95.3%.

Moreover, the historical commits of contributors were sourced from the third-party WoC database, resulting in a total of over 2.2 million commits. It is possible that tool instability during the data retrieval process could have caused some commits to be missed, potentially contributing to minor data omissions. However, considering the substantial number of commits collected, this limitation is unlikely to significantly affect the overall results.

Additionally, some GitHub contributors may have provided incomplete or inaccurate personal information, such as names or locations, which could introduce errors into the demographic data inferred from these attributes. While we employed gender inference software that is widely used and validated in academic research, the accuracy of such tools is inherently limited by the quality of input data. Although we aimed to cross-validate inferred demographic information using supplementary data sources such as social media profiles, the absence of such data for most contributors constrained our ability to do so. As a result, there is a possibility of misclassification in gender and regional attributes, which may affect the reliability of demographic analysis and introduce bias into region-specific or gender-specific findings. Furthermore, gender inference tools typically rely on name-based heuristics and often reflect binary gender classifications. These methods may not accurately capture non-binary, transgender, or gender-diverse identities, which presents an additional limitation to the inclusiveness and representativeness of our demographic analysis. To mitigate these limitations, we excluded contributors with ambiguous or missing name or region data and included only those with a gender inference confidence score above 0.7 in gender-related analysis [47]. While this approach improves reliability, we acknowledged that some degree of misclassification may remain.

5.3 Construct

In this study, we focused exclusively on commit activity to assess contributor participation in OSS and OSS4SG projects. While commits are a critical measure of software development, as they reflect tangible updates to the codebase [30], this metric alone does not capture the full range of contributions that shape OSS ecosystems. Other forms of activity, such as issue reporting, pull requests, code reviews, and discussion, also play significant roles in the development of OSS projects. By omitting these contributions, our analysis may present an incomplete picture of the overall dynamics. We acknowledge this limitation and recommend that future research incorporate a broader set of contribution types to provide a more comprehensive understanding of OSS4SG dynamics. To potentially mitigate the limitation, we collected contributors' full historical commits to provide a long-term view of participation that can help reveal deeper insights into their dynamics and impact. Moreover, this approach aligns with prior research that only employed commit frequency and volume as key indicators of contributor activity and participation in OSS development (e.g., [31, 37]). Finally, we defined "OSS4SG starter" based on the timing of observable commits, and it should be interpreted as early exposure to OSS4SG rather than a direct measure of intrinsic motivation, since we cannot fully disentangle voluntary contributions from those made within course or institutional programs.

In addition, when analyzing their long-term commit contributions, various social events, such as the COVID-19 pandemic, might influence their participation in OSS4SG. To mitigate this potential confounding factor, we evaluated the proportion of contributions to OSS4SG before and after the onset of COVID-19 and observed no significant differences. Therefore, the impact of COVID-19 on contributors' participation in OSS4SG appears to be minimal. However, we acknowledge that other localized factors across different regions may also have influenced contribution patterns.

6 Discussion

6.1 Challenges to Participation and Contribution in OSS4SG

Overall, contributors' engagement in OSS4SG remains consistently low throughout their contribution lifespan when compared to general OSS participation. Prior research suggests that developers often encounter difficulties in identifying new OSS4SG projects to contribute to [24], which may help explain part of the persistently limited involvement. Improving the visibility and accessibility of OSS4SG projects is therefore important and could help attract more contributors. For example, platforms could introduce standardized tags (e.g., #SocialGood, #CivicTech) or highlight OSS4SG projects in onboarding flows and issue recommendation systems (e.g., "Good First Issue-Social Good"). However, our findings show that even when contributors join OSS4SG, they exhibit lower overall participation and shorter active periods than in general OSS. This suggests that attracting alone is insufficient; effective strategies must also focus on retention and sustainability, ensuring that those who enter through OSS4SG remain engaged in the long term.

Interestingly, although these contributors engage in fewer OSS4SG projects, they tend to distribute their contributions more evenly over time in OSS4SG, compared to the short bursts of activity often seen in general OSS. One possible explanation for this pattern lies in social exchange theory [9], which suggests that contributors may view OSS4SG as a "high-cost, high-reward" environment. In such a context, their sustained but limited participation may reflect a rational trade-off, where intrinsic rewards, such as social impact or ethical fulfillment, outweigh more conventional extrinsic incentives like career advancement. However, their overall active period in OSS4SG is shorter than in general OSS. This may be due to a decline in perceived rewards over time, such as frustration from a lack of visible impact or emotional fatigue, which can lead to early departure. In contrast, general OSS often offers more enduring extrinsic motivators (e.g., career or reputation gains) that help maintain long-term participation. The relatively short windows of activity therefore raise concerns regarding OSS4SG project continuity and contributor retention.

Building on Clary and Snyder's functional approach to volunteer motivation [8], this pattern may be interpreted as evidence that many OSS4SG contributions are driven by values and understanding functions, rather than career or social functions. Contributors who are mainly values-driven may engage regularly but only for a limited period, especially when they do not see complementary benefits such as skill signaling, recognition, or stronger social ties. Moreover, from a communities-of-practice perspective [32, 58], the relative lack of institutional infrastructure in many OSS4SG communities, such as mentorship programs, clear role progression, or visible pathways from "first-time contributor" to core member, limits contributors' ability to move beyond peripheral participation, further constraining long-term engagement and differentiating OSS4SG from more established OSS communities. In this view, mentorship and structured newcomer pathways are not just practical supports but mechanisms of legitimate peripheral participation that integrate newcomers into the community.

6.2 Advancing Contribution Potential in OSS4SG

To strengthen contributions to OSS4SG, educational contexts offer promising entry points. Our findings show that contributors who begin with OSS4SG maintain a stronger tendency to engage with OSS4SG across their contribution histories. Integrating OSS4SG into curricula can therefore engage newcomers to open source early in their journey and channel their initial enthusiasm toward meaningful contributions to the OSS4SG community, while simultaneously equipping them with skills that transfer to general OSS. At the same time, prior work suggests that some developers hesitate to contribute to OSS4SG projects due to concerns about their potentially shorter lifespans [24], such as COVID-19-related efforts. In contrast, our analysis of seed projects indicates that, on average, the lifespan of OSS4SG projects is comparable to that of general OSS projects, a finding that should bolster confidence and encourage developers to contribute without assuming that OSS4SG projects are inherently short-lived.

To support more sustained engagement and reduce early departure, OSS4SG projects should adopt incentive strategies that align with contributors' motivations and aspirations, not only by recognizing meaningful contributions but also by connecting participation to personal and professional growth, as is commonly observed in general OSS. Building on the values-driven, episodic participation patterns identified in Sec. 6.1, Clary and Snyder's functional approach to volunteering [8] suggests that retention improves when opportunities simultaneously satisfy multiple motivational functions. For OSS4SG, this implies designing contribution opportunities that combine *values* (clear articulation of social impact), *understanding* (learning about domain issues and technologies), *career* (recognizable skills, credentials, or portfolio artifacts), and *social* functions (belonging to a community). In practice, scoping issues at different levels of complexity and time commitment allows projects to speak to different motivational functions: short, self-contained tasks can primarily satisfy values and understanding motives, whereas more complex, longitudinal tasks can engage career and social functions by building expertise and relationships. For example, issue descriptions can foreground both the social impact (e.g., "improves maternal health data for rural clinics") and the relevant technical stack, while contribution records can be made legible as portfolio items or credentials that signal expertise to employers. Since career development is a key motivator for many contributors on GitHub, OSS4SG organizers can further position their projects as valuable opportunities for skill building by structuring tasks to match contributors' career goals (e.g., tagging issues with in-demand technologies, simulating real-world workflows) and by partnering with employers or educational programs that explicitly value socially impactful open-source experience.

Our analysis (see Sec. 4.2) also suggests that strategies should avoid positioning OSS4SG as separate from the broader OSS ecosystem. Contributors who start in OSS4SG but remain isolated within it show limited productivity, whereas those who expand to general OSS maintain both high activity levels and a substantial share of OSS4SG engagement. This suggests that OSS4SG organizers should actively facilitate contributors' growth into the broader OSS community through cross-project recommendations, skill-development pathways that span OSS4SG and general OSS, and recognition systems that value both social impact and technical contributions. Social identity theory [23] helps explain these trajectories: contributors who begin in OSS4SG are likely to develop an early identity as "social impact" or "civic tech" developers, which helps sustain a relatively high proportion of OSS4SG activity even as their absolute contribution levels decline. However, when OSS4SG communities remain loosely connected to the larger OSS ecosystem, this identity is not reinforced by visible status, career progression, or community recognition. From a communities-of-practice perspective [32, 58], OSS4SG projects can thus be seen as entry points into a wider OSS practice: retention improves when contributors can deepen their participation across projects, acquire new roles, and have their identity as "social good developers" validated within the larger OSS community.

6.3 Variability by Region in OSS4SG

Contributors from developing regions, such as Africa, show a higher likelihood of participating in OSS4SG and of beginning their open-source journey with OSS4SG, yet the absolute number of contributors from these regions remains low. This combination suggests both a meaningful alignment with OSS4SG missions among those who do participate and persistent barriers to entry and sustained involvement. From a social identity perspective [23], projects that address locally salient issues (e.g., health, education, or infrastructure in under-resourced areas) may resonate strongly with contributors' identities and community ties, helping explain why those who do participate are disproportionately likely to choose OSS4SG. Yet if project communication, leadership, and recognition practices are dominated by norms and actors from more developed regions, contributors from developing regions may not fully experience a sense of belonging or visibility in the broader OSS4SG community.

These patterns highlight the need for targeted outreach and support efforts that both lower structural barriers and affirm contributors' identities in these regions. Potential interventions include translating OSS4SG documentation and onboarding resources into local languages, establishing region-specific recognition programs to highlight impactful contributions, and providing support through microgrants, digital badges, or sponsored events. Platforms such as GitHub can play an important role in facilitating these efforts by surfacing OSS4SG projects that are locally relevant and by making regionally targeted calls for participation more visible. Extending the communities-of-practice perspective [32, 58], such interventions can help move contributors from peripheral to more central participation within their regional contexts by clarifying pathways into the community, building local mentorship networks, and ensuring that contributions from developing regions are recognized and valued rather than remaining invisible in global metrics and discussions.

Educators in developing regions can also serve as key enablers in ways that complement the global educational strategies discussed in Sec. 6.2. By supporting students in overcoming infrastructural and technical hurdles when contributing to OSS4SG (e.g., unreliable connectivity, limited access to development tools), they can align coursework and capstone or graduation projects with locally relevant social issues and regional OSS4SG initiatives. Integrating OSS4SG projects into curricula in this way not only addresses regional social needs but also provides students who may have fewer opportunities for internships or industry experience with concrete opportunities to build skills, develop an identity as "social good" or "civic tech" developers in their own communities, and begin trajectories of participation that can extend into both OSS4SG and the broader open-source ecosystem, thereby fostering more geographically inclusive sustainability.

6.4 Variability by Gender in OSS4SG

Our findings also reveal a notable trend that women show a higher propensity to participate in OSS4SG projects. However, despite this greater relative participation, the overall number of their commit contributions remains lower than that of men, primarily due to the smaller number of women on GitHub. This disparity points to an opportunity for intervention aimed at strengthening the development and sustainability of OSS4SG. Specifically, implementing targeted initiatives on GitHub to attract more women to open source, particularly OSS4SG, and support their sustained and active participation. For example, given that women are more inclined to start their open-source contributions through OSS4SG, we can use OSS4SG as an effective tool for attracting and engaging more women in open-source development. GitHub could also spotlight the work of women in OSS4SG through monthly showcases and actively support the formation of women-focused OSS4SG communities on the platform. In addition, using generative AI tools to provide contextually relevant analysis and prompts could support the promotion of diversity. For example, a prompt might present: *"Currently, only 10% of contributors to this project identify as women. Would you like to help increase diverse participation?"* Such initiatives may not only recognize and amplify the contributions of women but also raise awareness about gender disparities within projects and encourage broader participation. Partnering with organizations such as AnitaB.org or other women-in-tech groups can help position OSS4SG as an inclusive and accessible entry point for women in open source. When engaging women, emphasizing the social impact and mission-driven goals of OSS4SG projects, especially how they can benefit women and their communities, may be particularly compelling and could enhance both initial interest and sustained participation.

7 Conclusion

OSS4SG, as a sub-community within the broader OSS community, has been attracting increasing attention. Several efforts have been made to enhance the visibility of OSS4SG. However, contributors' contribution patterns within OSS4SG remain poorly understood, even though these insights are critical for its long-term success and sustainability. In our study, we analyzed over 2.2 million commits from 5,860 contributors to explore how they contribute to OSS4SG and the factors that may influence their contributions. Our findings indicate that contributors show lower contribution intensity and shorter active duration in OSS4SG, yet their OSS4SG contributions are relatively more regular compared to OSS. Furthermore, contributors who begin with OSS4SG projects typically sustain higher participation levels in such projects throughout their contribution lifecycle. However, contributing to OSS4SG does not necessarily lead to sustained or increased contributions to these projects over time. Additionally, contributors from developing regions and women are more likely to begin their open-source journeys with OSS4SG and demonstrate a stronger tendency to remain engaged in these projects. Based on our findings, OSS4SG should be recognized as a pathway that attracts underrepresented groups, but one that requires targeted retention strategies to avoid early departure. Additionally, we propose targeted strategies to amplify the social impact of OSS4SG by engaging a diverse pool of contributors and supporting the long-term success of these projects. Beyond OSS4SG, our findings may also provide valuable insights for other technology-for-social-good initiatives, which often encounter similar challenges of inclusivity and sustained engagement. Effectively addressing these challenges is crucial to harnessing the potential of open-source innovation for tackling broader societal issues.

8 Data Availability

We made our data publicly available here.

References

[1] Oliver Alexy, Joachim Henkel, and Martin W Wallin. 2013. From closed to open: Job role changes, individual predispositions, and the adoption of commercial open source software development. *Research Policy* 42, 8 (2013), 1325–1340.

[2] 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.

[3] Amiangshu Bosu and Jeffrey C Carver. 2014. Impact of developer reputation on code review outcomes in oss projects: An empirical investigation. In *Proceedings of the 8th ACM/IEEE international symposium on empirical software engineering and measurement.* 1–10.

[4] Robert J Cabin and Randall J Mitchell. 2000. To Bonferroni or not to Bonferroni: when and how are the questions. *Bulletin of the ecological society of America* 81, 3 (2000), 246–248.

[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] Lars Carlsen and Rainer Bruggemann. 2022. The 17 United Nations' sustainable development goals: A status by 2020. *International Journal of Sustainable Development & World Ecology* 29, 3 (2022), 219–229.

[7] Antonio Cerone. 2012. Learning and activity patterns in OSS communities and their impact on software quality. *Electronic Communications of the EASST* 48 (2012).

[8] E Gil Clary and Mark Snyder. 1999. The motivations to volunteer: Theoretical and practical considerations. *Current directions in psychological science* 8, 5 (1999), 156–159.

[9] Russell Cropanzano and Marie S Mitchell. 2005. Social exchange theory: An interdisciplinary review. *Journal of management* 31, 6 (2005), 874–900.

[10] Sherae Daniel, Ritu Agarwal, and Katherine J Stewart. 2013. The effects of diversity in global, distributed collectives: A study of open source project success. *Information Systems Research* 24, 2 (2013), 312–333.

[11] Kareem Abbas Dawood, Khaironi Yatim Sharif, AA Zaidan, Abdul Azim Abd Ghani, Hazura Binti Zulzalil, and BB Zaidan. 2019. Mapping and analysis of open source software (OSS) usability for sustainable OSS product. *IEEE Access* 7 (2019), 65913–65933.

[12] Luis Felipe Dias, Igor Steinmacher, and Gustavo Pinto. 2018. Who drives company-owned OSS projects: internal or external members? *Journal of the Brazilian Computer Society* 24, 1 (2018), 16.

[13] Ikram El Asri and Noureddine Kerzazi. 2019. Where are females in oss projects? socio technical interactions. In *Collaborative Networks and Digital Transformation: 20th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2019, Turin, Italy, September 23–25, 2019, Proceedings 20.* Springer, 308–319.

[14] Vijaya Kumar Eluri, Thomas A Mazzuchi, and Shahram Sarkani. 2021. Predicting long-time contributors for GitHub projects using machine learning. *Information and Software Technology* 138 (2021), 106616.

[15] Zihan Fang, Madeline Endres, Thomas Zimmermann, Denae Ford, Westley Weimer, Kevin Leach, and Yu Huang. 2023. A Four-Year Study of Student Contributions to OSS vs. OSS4SG with a Lightweight Intervention. (2023).

[16] Denae Ford, Nischal Shrestha, and Thomas Zimmermann. 2022. ReBOC: Recommending Bespoke Open Source Software Projects to Contributors. In *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC).* 1–5. doi:10.1109/VL/HCC53370.2022.9833113

[17] Todd Michael Franke, Timothy Ho, and Christina A Christie. 2012. The chi-square test: Often used and more often misinterpreted. *American journal of evaluation* 33, 3 (2012), 448–458.

[18] Hana Frluckaj, Laura Dabbish, David Gray Widder, Huilian Sophie Qiu, and James D Herbsleb. 2022. Gender and participation in open source software development. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.

[19] Marco Gerosa, Igor Wiese, Bianca Trinkenreich, Georg Link, Gregorio Robles, Christoph Treude, Igor Steinmacher, and Anita Sarma. 2021. The shifting sands of motivation: Revisiting what drives contributors in open source. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE).* IEEE, 1046–1058.

[20] Baskaran Govindaras, Tuan Sau Wern, Sharangeet Kaur, Idris Akmal Haslin, and R Kanesaraj Ramasamy. 2023. Sustainable environment to prevent burnout and attrition in project management. *Sustainability* 15, 3 (2023), 2364.

[21] Mariam Guizani, Amreeta Chatterjee, Bianca Trinkenreich, Mary Evelyn May, Geraldine J Noa-Guevara, Liam James Russell, Griselda G Cuevas Zambrano, Daniel Izquierdo-Cortazar, Igor Steinmacher, Marco A Gerosa, et al. 2021. The long road ahead: Ongoing challenges in contributing to large oss organizations and what to do. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–30.

[22] Mariam Guizani, Thomas Zimmermann, Anita Sarma, and Denae Ford. 2022. Attracting and retaining oss contributors with a maintainer dashboard. In *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society.* 36–40.

[23] Michael A Hogg. 2016. Social identity theory. In *Understanding peace and conflict through social identity theory: Contemporary global perspectives*. Springer, 3–17.

[24] Yu Huang, Denae Ford, and Thomas Zimmermann. 2021. Leaving My Fingerprints: Motivations and Challenges of Contributing to OSS for Social Good. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. 1020–1032. doi:10.1109/ICSE43902.2021.00096

[25] Md Hasan Ibrahim, Mohammed Sayagh, and Ahmed E Hassan. 2021. A study of how Docker Compose is used to compose multi-component systems. *Empirical Software Engineering* 26 (2021), 1–27.

[26] Aftab Iqbal. 2014. Understanding contributor to developer turnover patterns in oss projects: a case study of apache projects. *International Scholarly Research Notices* 2014, 1 (2014), 535724.

[27] Javier Luis Cánovas Izquierdo and Jordi Cabot. 2018. The role of foundations in open source projects. In *Proceedings of the 40th international conference on software engineering: software engineering in society*. 3–12.

[28] Jack Jamieson, Naomi Yamashita, and Eureka Foong. 2024. Predicting open source contributor turnover from value-related discussions: An analysis of GitHub issues. In *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*. 1–13.

[29] Eirini Kalliamvakou, Georgios Gousios, Kelly Blincoe, Leif Singer, Daniel M German, and Daniela Damian. 2014. The promises and perils of mining github. In *Proceedings of the 11th working conference on mining software repositories*. 92–101.

[30] Carsten Kolassa, Dirk Riehle, and Michel A Salim. 2013. The empirical commit frequency distribution of open source projects. In *Proceedings of the 9th international symposium on open collaboration*. 1–8.

[31] Carsten Kolassa, Dirk Riehle, and Michel A. Salim. 2013. The empirical commit frequency distribution of open source projects. In *Proceedings of the 9th International Symposium on Open Collaboration* (Hong Kong, China) *(WikiSym '13)*. Association for Computing Machinery, New York, NY, USA, Article 18, 8 pages. doi:10.1145/2491055.2491073

[32] Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge university press.

[33] Sihai Lin, Yutao Ma, and Jianxun Chen. 2013. Empirical Evidence on Developer's Commit Activity for Open-Source Software Projects.. In *Seke*, Vol. 13. 455–460.

[34] Yuxing Ma, Chris Bogart, Sadika Amreen, Russell Zaretzki, and Audris Mockus. 2019. World of code: an infrastructure for mining the universe of open source VCS data. In *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*. IEEE, 143–154.

[35] Yuxing Ma, Tapajit Dey, Chris Bogart, Sadika Amreen, Marat Valiev, Adam Tutko, David Kennard, Russell Zaretzki, and Audris Mockus. 2021. World of code: enabling a research workflow for mining and analyzing the universe of open source VCS data. *Empirical Software Engineering* 26 (2021), 1–42.

[36] Yutao Ma, Yang Wu, and Youwei Xu. 2014. Dynamics of open-source software developer's commit behavior: An empirical investigation of Subversion. In *Proceedings of the 29th annual ACM symposium on applied computing*. 1171–1173.

[37] Yutao Ma, Yang Wu, and Youwei Xu. 2014. Dynamics of open-source software developer's commit behavior: an empirical investigation of subversion. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing* (Gyeongju, Republic of Korea) *(SAC '14)*. Association for Computing Machinery, New York, NY, USA, 1171–1173. doi:10.1145/2554850.2555079

[38] Patrick E McKnight and Julius Najab. 2010. Mann-Whitney U Test. *The Corsini encyclopedia of psychology* (2010), 1–1.

[39] Régis Meissonierm, Isabelle Bourdon, Serge Amabile, and Stephane Boudrandi. 2012. Toward an enacted approach to understanding oss developer's motivations. *International Journal of Technology and Human Interaction (IJTHI)* 8, 1 (2012), 38–54.

[40] 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.

[41] Audris Mockus, Roy T Fielding, and James D Herbsleb. 2002. Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 11, 3 (2002), 309–346.

[42] Kumiyo Nakakoji, Yasuhiro Yamamoto, Yoshiyuki Nishinaka, Kouichi Kishida, and Yunwen Ye. 2002. Evolution patterns of open-source software systems and communities. In *Proceedings of the international workshop on Principles of software evolution*. 76–85.

[43] Saya Onoue, Hideaki Hata, and Ken-ichi Matsumoto. 2013. A Study of the Characteristics of Developers' Activities in GitHub. In *2013 20th Asia-Pacific Software Engineering Conference (APSEC)*, Vol. 2. 7–12. doi:10.1109/APSEC.2013.104

[44] 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.

[45] Raphael Pham, Leif Singer, and Kurt Schneider. 2013. Building test suites in social coding sites by leveraging drive-by commits. In *2013 35th International Conference on Software Engineering (ICSE)*. IEEE, 1209–1212.

[46] 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* 48, 9 (2021), 3394–3409.

[47] Huilian Sophie Qiu, Zihe H Zhao, Tielin Katy Yu, Justin Wang, Alexander Ma, Hongbo Fang, Laura Dabbish, and Bogdan Vasilescu. 2023. Gender representation among contributors to open-source infrastructure: an analysis of 20 package manager ecosystems. In *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 180–187.

[48] Ayushi Rastogi. 2016. Do biases related to geographical location influence work-related decisions in github?. In *Proceedings of the 38th international conference on software engineering companion*. 665–667.

[49] Ayushi Rastogi, Nachiappan Nagappan, and Georgios Gousios. 2016. *Geographical bias in GitHub: Perceptions and reality*. Technical Report.

[50] Ayushi Rastogi, Nachiappan Nagappan, Georgios Gousios, and André van der Hoek. 2018. Relationship between geographical location and evaluation of developer contributions in github. In *Proceedings of the 12th ACM/IEEE international symposium on empirical software engineering and measurement*. 1–8.

[51] Lucía Santamaría and Helena Mihaljević. 2018. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science* 4 (2018), e156.

[52] 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.

[53] Josh Terrell, Andrew Kofink, Justin Middleton, Clarissa Rainear, Emerson Murphy-Hill, Chris Parnin, and Jon Stallings. 2017. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science* 3 (2017), e111.

[54] Bianca Trinkenreich, Mariam Guizani, Igor Wiese, Anita Sarma, and Igor Steinmacher. 2020. Hidden figures: Roles and pathways of successful oss contributors. *Proceedings of the ACM on human-computer interaction* 4, CSCW2 (2020), 1–22.

[55] Georg Von Krogh, Stefan Haefliger, Sebastian Spaeth, and Martin W Wallin. 2012. Carrots and rainbows: Motivation and social practice in open source software development. *MIS quarterly* (2012), 649–676.

[56] Johannes Wachs, Mariusz Nitecki, William Schueller, and Axel Polleres. 2022. The geography of open source software: evidence from github. *Technological Forecasting and Social Change* 176 (2022), 121478.

[57] Zhendong Wang, Yang Feng, Yi Wang, James A Jones, and David Redmiles. 2020. Unveiling elite developers' activities in open source projects. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 29, 3 (2020), 1–35.

[58] Etienne Wenger. 1999. *Communities of practice: Learning, meaning, and identity*. Cambridge university press.

[59] Robert F Woolson. 2005. Wilcoxon signed-rank test. *Encyclopedia of biostatistics* 8 (2005).

[60] Qi Xuan and Vladimir Filkov. 2014. Building it together: Synchronous development in OSS. In *Proceedings of the 36th International Conference on Software Engineering*. 222–233.

[61] Yunwen Ye and Kouichi Kishida. 2003. Toward an understanding of the motivation of open source software developers. In *25th International Conference on Software Engineering, 2003. Proceedings*. IEEE, 419–429.

[62] Likang Yin, Mahasweta Chakraborti, Yibo Yan, Charles Schweik, Seth Frey, and Vladimir Filkov. 2022. Open source software sustainability: Combining institutional analysis and socio-technical networks. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–23.

[63] Yang Yue, Yi Wang, and David Redmiles. 2023. Off to a Good Start: Dynamic Contribution Patterns and Technical Success in an OSS Newcomer's Early Career. *IEEE Transactions on Software Engineering* 49, 2 (2023), 529–548. doi:10.1109/TSE.2022.3156071

[64] 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.

Received May 2025; revised November 2025; accepted December 2025