Diversity and Inclusion in Open Source Software (OSS) Projects: Where Do We Stand?

Amiangshu Bosu
Department of Computer Science,
Wayne State University
Detroit, MI, USA
abosu@wayne.edu

Kazi Zakia Sultana
Department of Computer Science,
Montclair State University
Montclair, NJ, USA
sultanak@montclair.edu

***Abstract*—*Background:* As the area of computing has thrived over the years, the participation of women in computing declined. Currently women represent less than 24% of the computing workforce and that number is declining. On the other hand, the ratios of women in Open Source Software (OSS) projects are even lower.**

Aims:* The primary objective of this study is *to determine the level of gender diversity among popular OSS projects and identify the presence of gender biases that may discourage females' participation.

***Method:* On this goal, we mined the code review repositories of ten popular OSS projects. We used a semi-automated approach followed by a manual validation to identify the genders of the active contributors.**

***Results:* Our results suggest that lack of gender diversity remains an ongoing issue among all the ten projects as each of the projects had less than 10% female developers. However, many of the projects also suffer from lack of inclusion of females to leadership positions. Although none of the projects suggest significant differences between male and female developers in terms of productivity based on three different measures, data from three out of the ten projects indicate technical biases against female developers with lower code acceptance rates as well as delayed feedback during code reviews. However, biases against females are not universal as majority of the projects do not discriminate against females. The two projects with the least ratios of female contributors as core developers showed the most biases against females.**

***Conclusion:* Based on our findings, we conclude that promoting and mentoring females to leadership positions may be an effective solution to foster gender diversity.**

***Index Terms*—gender issues, diversity, open source projects, inclusion, discrimination**

I. Introduction

According to the US Bureau of Labor Statistics, in 1995, 37% of the computer scientists were females [19]; twenty two years later, in 2017, that number dropped to only 24% [2]. On the other hand, the ratios of women in Open Source Software (OSS) projects are even lower as recent surveys found, only 5.2% females in Apache Software Foundation [42] and 9.9% females in Linux kernel [11]. Recent studies investigating OSS projects hosted on Github also found less than 10% female contributors [48], [44], [27]. Although gender diversity positively impacts the productivity of OSS projects [47], [35], most of the contemporary OSS projects lack diversity. Therefore, our primary research objective is to promote diversity among OSS communities. To achieve this long-term goal, this study is a preliminary step that aims *to determine the level of gender diversity among popular OSS projects and identify the presence of gender biases that may discourage females' participation.*

Several recent Software Engineering (SE) studies have also investigated diversity issues and gender biases among OSS projects [44], [37], [20], [33]. Most notably, Terrell *et* al. investigated the acceptance of pull requests among Github contributors and found women's pull requests having significant more acceptance rate than men's (women: 78.7% to men: 74.6%); however, the acceptance rate of women's pull requests drops to only 58%, when women's gender is identifiable, suggesting possible biases against women [44]. Another recent article [37] investigated the factors behind the very low ratios of women among the OSS projects and suggested couple of factors to blame. First, unwelcoming behaviors against women in tech industries [21], [34], [29] negatively impact the confidence levels of female developers, which demotivates them from joining OSS projects. Second, OSS projects have been guilty of fostering a 'toxic culture' for women, where inflammatory talk and aggressive posturing is acceptable within the norm of the community [33]. Since female-led OSS projects, which may have cultivated female-friendly cultures, are hardly found, many capable women stay away from OSS projects.

However, OSS projects are crucial as today's web infrastructure is dominated by OSS (e.g., Linux kernel, Chromium, Apache, and OpenSSL). Recognizing the importance of OSS, all the large computing organizations (e.g., Microsoft, Google, and Facebook) are actively contributing to OSS projects [5]. As SE researchers have already found positive impacts of gender diversity on software development team productivity [47], [35], we believe that OSS projects would also be more productive if those can attract and recruit more women, who comprise almost half of the population. Unfortunately, the even lower ratios of women in OSS projects than the computing industry suggest those projects as more female-unfriendly and uninteresting to women. Moreover, this lack of diversities among OSS projects could make the computing industry's ingrained imbalances worse, since a recent survey [20] found almost half of the respondents agreeing that their OSS contributions played an active role in landing their current computing jobs. Therefore, we believe the lack of diversity among OSS projects needs to be investigated from various perspectives. Hence, this study aims to add another perspective to analyze this issue.

This study is significantly motivated by Terrell *et* al.'s findings. However, our additional investigations as summarized following differentiate this study from Terrell *et* al.'s. First, while Terrell *et* al.'s study investigated all Github contributors from a large number of projects together, we investigate ten popular OSS projects separately, since we hypothesize that some OSS projects may be exhibiting more gender biases than others. Second, since prior research suggests women in computing lack access to creative technical roles [6], we investigate whether these projects are inclusive to female developers. Third, since many anecdotes questioned the capabilities of women as software developers [17], we investigate whether there is a significant difference in productivity based on the gender of a contributor. Finally, we explore the impacts of biases against women among the projects under study.

To investigate the goals of this study, we mined and examined the code review repositories of ten popular OSS projects. We selected code reviews as those facilitate direct communications between developers as a Mozilla developer stated, "In the software world, code review is a primary tool that we use to communicate, to assign value to our work, and to establish the pecking order at work in our industry. [30]" Based on the approach adopted in Terrell *et* al.'s study [44], we developed a semi-automated methodology followed by a manual validation using social networks (i.e., LinkedIn, Google Plus, Facebook, Github, and Twitter) to identify the genders of the 'Non-casual developer'[1]. In summary, the primary contributions of this study are:

- An investigation of the ratios of women as contributors as well as core developers among ten popular OSS projects.
- A comparison of the productivity of female developers against their male peers.
- An investigation of technical biases for / against women in ten OSS projects.
- A preliminary investigation of the impacts of biases, if exists, on the participation of female developers.

The remainder of the paper is organized as follows. Section II provides background about gender bias in computing and code reviews. Section III introduces the research questions of this study. Section IV describes our research methodology. Section V presents the results of this study. Section VI discusses the implications of the results. Section VII describes the threats to validity of our findings. Finally, Section VIII provides some directions for future work and concludes the paper.

II. Background

This section provides a brief background on three areas to understand our study design: i) gender bias, ii) prior research on gender issues in OSS, and iii) code reviews.

A. Gender Bias

Bias is a person or entity's inclination towards or prejudice against one thing, person, or group compared with another usually in a way that's considered to be unfair [3]. Similarly, a person or entity's bias primarily based on another person's gender is 'gender bias'. There are primarily two types of biases:

1) Conscious bias (or explicit bias): Conscious bias is to be aware, intentional and responsive. It is a tendency, trend, inclination, feeling or opinion, which is particularly preconceived and certainly unreasonable. Following are few examples of conscious biases that women often encounter in OSS projects [1].

- Some male developers often speak or act as if women don't exist in OSS community at all.
- Some male developers often describe a dichotomy of "geeks" and "women" in such a way that the idea of "women geeks" is impossible.
- Some developers treat male-led OSS projects differently compared to women-led ones. For example, a man leading a tiny FLOSS project with few users asked to merge a woman led project which is even larger and consists of many users than his.
- Some perceive code contributed by a woman as beginner-quality, or does not consider a female-led project as a real project.
- Some male developers make negative comments about a woman's OSS code without her knowledge leaving no chance for the woman to respond.

2) Unconscious bias (or implicit bias): Unconscious biases are social stereotypes about certain groups of people that individuals form outside their own conscious awareness. Everyone holds unconscious beliefs about various social and identity groups, and these biases stem from one's tendency to organize social worlds by categorizing. Unconscious bias is far more prevalent than conscious prejudice and often incompatible with one's conscious values. Certain scenarios can activate unconscious attitudes and beliefs. There is evidence of widespread prejudice against women and girls from decades of psychological research [4].

Following are few examples of unconscious biases that women often encounter in OSS projects.

- In our society, there is a trend of belief that women have natural instinct to do better or worse at certain tasks, or having different interests. In the OSS community, this norm often plays out as an excuse to describe why the proportion of women in the field is so low and the actual reasons are not investigated at all.
- People often strain to argue that women are 'biologically' and 'culturally' good and trained at certain tasks such as communication, visual design, documentation. Therefore, women in some OSS projects are often forced to do certain activities like documentation, tutorial or UI work even though they do not feel interest in them.

[1] A developer who has submitted at least five code changes to a project under our study.

- Due to the paucity of women in OSS, the community often behaves it as an all-male environment. For example, desktop wallpapers, advertisements, and conference presentations are often decorated with some graphics and texts that are uncomfortable to the women counterparts.

B. Prior works on Gender Issues in OSS

Earlier studies in software engineering focused on the participation of women in FLOSS projects and reported extremely lower (i.e., between 2% to 5%) ratio of female contributors [22]. The reason behind the lower participation of female is attributed to the social and cultural arrangements of OSS projects, which actively exclude female contributors [33]. Due to several initiatives from few of the OSS projects (i.e., Debian, and GNOME) to attract female contributors, the number of female contributors has improved to around 10% in some projects [40], however the number is still less than half of the industry average (i.e., 24%). Similar to other domains[15], [23], SE researchers have also observed the positive impacts of gender diversity on software development team productivity [47], [35]. Yet, studies also found various types of discrimination against women. For example, women in computing organizations are often assigned menial tasks, while similar male colleagues are given ‘choice’ projects [18], often do not get opportunities at management positions, and earn lower salaries compared to men [36]. Another survey found 72% women perceiving themselves as victims of unfair evaluations due to their genders [24]. As a result, female developers become increasingly pessimistic about career opportunities as their tenure progresses [25]. Even in OSS projects, where most of the participants are volunteers, women perform majority other types of contributions than coding, while men mostly contribute with code [40].

C. Code Review

Peer code review is a software engineering practice, where a developer sends his/her code to a peer to identify possible defects before merging to the project codebase. Compared with the traditional heavy-weight inspection process, peer code review is more informal, tool-based, and used regularly in practice [7]. To make peer code reviews more efficient, teams use automated support tools such as Gerrit[2], Phabricator[3], and ReviewBoard[4]. A tool-based code review process starts when an author creates a *patchset* (i.e. all files added or modified in a single revision) along with a description of the changes and submits that information to a code review tool. To facilitate reviews, code review tools highlight the changes between two revisions in a side-by-side display. Both the reviewers and the author can insert comments pointing out issues, suggesting improvements, or clarifying the changes. After the review, the author may upload a new patch-set addressing the review comments and initiate a new review iteration. This review cycle repeats until either the reviewers approve the change or the author abandons it. Code review tools capture the interactions (a.k.a. review comments) between the author and a reviewer to facilitate *post-hoc* analyses.

III. Research Questions

The primary objective of this study is *to determine the level of gender diversity among popular OSS projects and identify the presence of gender biases that may discourage women's participation.* Following subsections introduce four specific research questions to investigate this objective.

A. Gender Diversity

The percentage of the active female developers is very low compared to male developers. While most of the studies report around 10% females, a recent study suggests even bleaker pictures with only 3% women [50]. We aim to estimate the gender diversities of the popular OSS projects. Hence our first research question is:

RQ1: *What are the percentages of female contributors in popular OSS projects?*

B. Gender Inclusion

Prior research found women in computing organizations were often assigned menial tasks (e.g., documentation, data labeling, and manual testing) [18] and often did not get opportunities at management positions [36]. Therefore, we seek to find out if the popular OSS projects are gender-inclusive with promoting females to leadership roles.

RQ2: *What percentage of leadership roles are occupied by females among the OSS projects?*

C. Productivity Comparison

"At a recent tech event, I overheard a side conversation about the lack of gender diversity in tech. The small group was discussing the fact that even though women make up about 30 percent of the workforce in tech, higher level engineering teams rarely have more than a few women. One of the participants in this conversation commented that this was because male developers are just generally more talented than female developers. No one in the group objected." [37]. Therefore, our third research question compares the productivity of female developers with the productivity of their male colleagues.

RQ3: *Are female developers of OSS projects significantly more/less productive than their male colleagues?*

D. Explicit biases for / against women

A study of more than 50 GitHub repositories [44] revealed that women's pull requests were approved more often than their male counterparts - nearly 3% more often. However, if their gender was known, female coders were .8% less likely to have their code accepted. Our next research question investigates if similar biases are prevalent among the popular OSS projects. We focus only on explicit biases, since we can define and measure various metrics to estimate those.

RQ4: *Is there any explicit bias for or against female developers in open source projects?*

[2]https://code.google.com/p/gerrit/
[3]http://phabricator.org/
[4]https://www.reviewboard.org/

IV. Research Methodology

The accurate identification of the gender of a contributor is not only essential but also challenging for this research. We developed our gender resolution approach for our recent study [38] investigating gender biases in OSS projects from another perspective. The design and analysis of that study is available in the proceedings of a recent conference. For that study, we resolved the genders of 3,570 non-casual contributors from six OSS projects. We augment that dataset by adding 973 developers from four additional OSS projects (i.e., Go, LibreOffice, Typo3, and Whamcloud). To facilitate the completeness of this paper, we repeat the gender resolutions steps here. In the following subsections, we describe our methods to collect data, resolve a developer's gender, and identify the set of core developers for each project.

A. Data Collection and Preparation

We used a Gerrit-Miner [13] tool to mine 683,865 code review requests from ten popular OSS projects and stored the data in a MySQL database. Our project selection was based on two criteria: i) an open source project actively using Gerrit; and ii) project contributors have performed at least 15,000 code reviews. Table I shows the list of projects. The projects selected in this study are well-known OSS projects and have been under active development.

A manual inspection of the comments posted by some accounts (e.g., 'Qt Sanity Bot' or 'BuildBot') suggested that those accounts were automated bots rather than humans. These accounts typically contain one of the following keywords: 'bot', 'auto', 'CI', 'Jenkins', 'integration', 'build', 'hook', 'recheck', 'travis', or 'verifier'. Because we wanted only code reviews from actual reviewers, we excluded these bot accounts after a manual inspection had confirmed that the interactions were automatically generated. Following a similar approach as Bird *et al.* [10], we used the Levenshtein distance between two names to identify similar names. If our manual reviews of the associated accounts suggested that those belong to the same person, we merged those to a single account.

In this study, we define a *'Non-casual developer'* as a developer who has submitted at least five code changes for his/her project. Since our gender resolution strategy is time consuming, we only considered the non-casual developers in each project for our subsequent analyses. Column 'non-casual devs.' in Table I shows the number of non-casual developers in each of the ten projects. Although, the number of non-casual developers may be as low as 38% of total developers (i.e., Android), they contributed more than 95% of total code changes in each of the ten projects. Therefore, an analysis of the non-casual developers should provide us adequate understandings of the communication and collaboration in an OSS project. Our data preparation steps generated a list of 4,543 non-casual developers (Table I: Non-casual devs.) from the ten projects.

TABLE I
PROJECT DEMOGRAPHICS

Project	**Sponsor**	**Domain**	**Requests mined***	**Total non-casual devs.**
Android	Google	Mobile OS	81,137	968
Chromium OS	Google	Desktop OS	153,523	1,007
Couchbase	Couchbase Inc.	NoSQL database	64,799	165
Go	Google	Programming	19,318	220
LibreOffice	Foundation	Office Suite	35,545	344
OmapZoom	Texas Instruments	Mobile Platform	35,973	425
oVirt	Redhat Inc.	Virtualization	73,523	220
Qt	Qt Company	UI framework	155,936	736
Typo3	Foundation	CMS	48,940	318
Whamcloud	Intel	HPC	15,171	140
		Total:	**683,865**	**4,543**

*Mined during September, 2017

B. Gender Resolution

We adopted a semi-automated gender resolution strategy using the genderComputer tool created by Vasilescu *et al.* [46] and modified by Terrell *et al.* [44] and followed the automated steps with manual validations using publicly available information on social networks.

The genderComputer tool uses a database of 221,854 first names from 204 countries around the world and classifies each name belonging to one of the following four categories:

1) *Male*- names given to males.
2) *Female*- names given to females.
3) *Unisex* - names given to both males and females.
4) *None*- no entry in the database.

Based on the names of our 4,543 non-casual developers from our ten projects, the genderComputer tool classified the contributors as following: 3,355 males, 400 females, 520 unisex, and 152 none. To ensure the accuracy of the identified genders, we adopted following five-step manual validation strategy for the 1,072 non-male (i.e., female, unisex, or none) contributors. We moved onto the next resolution step only if all the previous steps failed. If all the five manual validation steps were unsuccessful, which occurred for only 47 contributors ($\approx$ 1%), we excluded a developer from our subsequent analyses. Although, our gender resolution steps leverage only publicly available information, we got our research method reviewed and approved by our Institutional Review Board (IRB).

1) Resolution using Gerrit Avatar: Gerrit allows a user to include his/her picture in his/her profile. In the first step, we look into the Gerrit avatar of a user to determine his/her gender. Figure 1 shows examples of two Gerrit avatars. The avatar on the left indicates a male contributor and the avatar on the right indicates a female. However, some users' Gerrit avatars were either empty or images that do not reveal their genders.

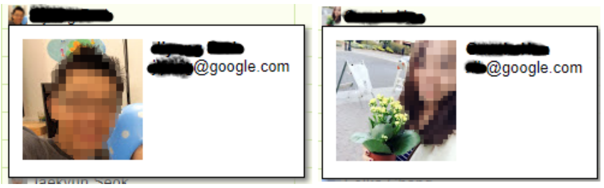

Fig. 1. Gender resolution using Gerrit profile pictures: 1) male (left), 2) female (right)

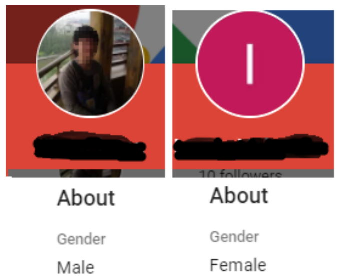

Fig. 2. Gender resolution using Google plus profile: 1) male (left), 2) female (right).

2) Resolution using Google plus: We searched the Google plus social network using an user's email address. Based on Google plus search policy, if a user has associated his/her email address with a profile, a search based on that email address returns only that particular profile. Since the gender information of a user on Google plus is public for the majority of the users, a positive match based on an email search potentially could help finding the user's gender (Figure 2).

3) Resolution using LinkedIn Profile: For these users, we searched LinkedIn, a professional social network, with his/her full name and company information. For example, if a user's name is '*Fang Liu*' and his/her email address is '*fliu@google.com*', we searched using the term '*Fang Liu + Google*'. If we found a positive match, we inspected the profile picture to determine his/her gender. However, if a user's profile picture was invisible to us, we looked into the recommendations that he/she has received. Any gender specific pronouns (i.e., 'he', 'she', 'his', or 'her') in the recommendations revealed the user's gender. Figure 3 shows examples of gender resolutions using the LinkedIn.

4) Resolution using Facebook: We used the same search term as used on the LinkedIn (*'full_name + company_name'*) on Facebook. If a positive match was found, we inspected the profile pictures as well as gender specific pronoun in the phrase ('To see what he/she shares..') to determine a user's gender. Figure 4 shows an example Facebook search that found a female contributor that worked for Google.

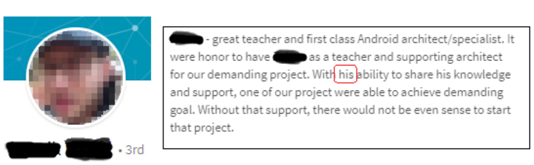

Fig. 3. Gender resolution: 1) using LinkedIn profile picture (left), 2) using received recommendations (right).

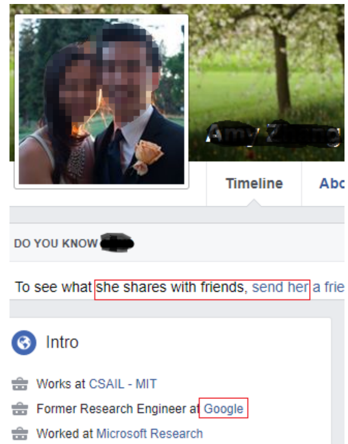

Fig. 4. Gender resolution: using *'full_name + company_name'* search on Facebook.

5) Resolution using Google Search: If all the first four steps failed ($< 3\%$ overall), we searched on Google using *'full_name + company_name'* to identify the profiles of a user on various other platforms (e.g., blog, presentation, video, Twitter, Github, and forums). If information obtained from those platforms suggest a positive match, we inspected pictures or referring pronouns on those platforms to identify the gender of a user.

C. Identification of Core developers

The structure of OSS development communities has often been described as a core-periphery structure [49], [32], with a small number of core developers and a larger set of peripheral developers. The core developers are those who have been involved with the OSS project for a relatively long time and have made significant contributions to guide the development and evolution of the project [49]. Due to their significant contributions and higher level of interactions, core developers are often considered leaders of OSS projects [49], [14]. Prior studies have used either social network analysis [14], [16] or number of code commit thresholds [32], [39] to identify the core developers of an OSS project. Since a recent study found both count-based and network-based classifications significantly agreeing with one another in ten different OSS projects [26], we use the code-commit-threshold based classification in this study. We classify the top 10% developers in terms of the number of code commits in a project as core developers for that project.

D. Metrics to estimate productivity and bias

We use three different measures (Table II) to estimate the productivity of a developer. To facilitate computation of those three measures, we compute the tenure of each developer by deducting the timestamp of his/her first submitted code commit from the timestamp of his/her most recent submitted code change in our dataset. We converted those tenures to number

TABLE II
METRICS TO ESTIMATE THE PRODUCTIVITY OF A DEVELOPER

Metrics	Definition
Code changes per month	On average, the number of code commits by a developer per month during his/her tenure with a project.
Code churn per month	On average, total lines of code churned (i.e., added /deleted /modified) by a developer per month during his/her tenure with a project.
Code reviews per month	On average, the number of code reviews a developer participated in per month during his/her tenure with a project.

TABLE III
METRICS TO ESTIMATE EXPLICIT BIASES

Metrics	Definition	Rationale
Acceptance rate	Percentage of submitted code reviews that are successfully merged.	The study by Terrell et al. [44] found females' code were less likely to be accepted.
First feedback interval	First Feedback Interval is the amount of time from the submission of a code review request in Gerrit until the first real review comment (i.e. not an automated comment).	Developers prefer reviewing code from peers that they know well or admire [12].
Review interval	Review Interval is the time from the beginning to the end of the review process.	Developers who are not well-connected with their peers are more likely to have higher review interval [14]
Code churn per comment	Code churn indicates the number of lines added, modified or deleted to a file from one version to another. Code churn per comment indicates number of code churn per code review comment.	A lower churn per comment would indicate higher scrutiny for a developer's code.

of months by dividing the number of seconds in 30 days and taking the ceiling.

We use four different metrics to compute explicit biases. Table III defines those metrics and provides a brief rationale behind their inclusions.

V. RESULTS

After successful resolution of the genders of more than 98% active contributors from the ten projects, we performed statistical analysis to answer our four research questions as described in the following subsections. Since our analyses compare two different groups of developers (i.e. Male and Female) from the same project multiple times based on different measures, we adjust the Statistical significance (p) levels using False Discovery Rate (FDR) [9] corrections. We use non-parametric tests for all comparisons, since the results of Shapiro-Wilk [41] tests suggest that the distributions of the productivity (Table II) and bias (Table III) metrics significantly differ from a normal distribution.

TABLE IV
GENDER DIVERSITY AND INCLUSION IN THE TEN PROJECTS

Project	# non-casual female dev.	% female	# core dev.	% female core dev.
Android	60	6.19%	258	3.87%
Chromium OS	88	8.74%	151	3.97%
Couchbase	16	9.69%	24	4.17%
Go	13	5.90%	90	7.77%
LibreOffice	28	8.14%	68	1.47%
OmapZoom	30	7.06%	60	10.00%
oVirt	21	9.54%	34	2.94%
Qt	23	3.10%	159	3.12%
Typo3	12	3.77%	73	4.10%
Whamcloud	11	7.85%	19	0.00%

A. Gender diversity

Table IV shows the ratio of female contributors in the ten projects in our study. All the ten projects have less than 10% non-casual female contributors. These ratios are similar to the ratios found in other recent studies [44], [48], [27], [42]. While the percentage of women in computing jobs has been steady at 24% over the last six years [2], during the same period the ratio of of women in top OSS projects were lower than half of that number.

Finding 1: *Consistent with prior results, the ratios of 'non-casual' female contributors are less than 10% among all ten projects.*

B. Gender Inclusion

For each of the project, we also identified the core contributors using the method described in Section IV-C. We hypothesize that if an OSS project is gender inclusive and provides female developers equal opportunities to become core contributors, we would expect close to proportional number of female developers as core developers. The column "% female core dev." in Table IV shows the percentage of female core developers in the ten OSS projects. For example, among the 258 core developers in the Android project only 10 developers (i.e., 3.87%) were females. Our results found only four (i.e., Go, Omapzoom, Qt, and Typo3) out of the ten projects with fair or better ratios of female core developers (i.e., by comparing "% female core dev." column with the "% female" column from Table IV).

Finding 2: *Six out of the ten projects have even lower ratios of females as core developers.*

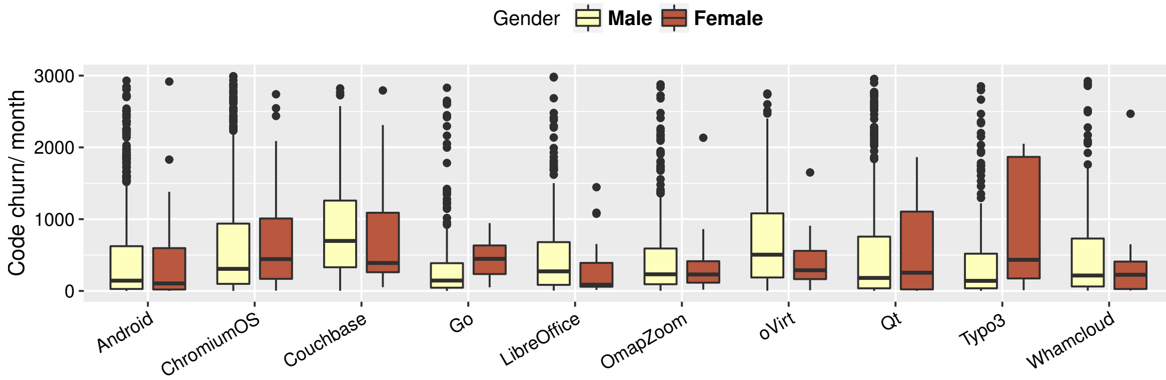

Fig. 5. Code churn per month (male vs. female)

TABLE V
PRODUCTIVITY: MALE VS. FEMALE

Project	**Code changes / month**			**Code churn / month**			**Code reviews /month**		
	Male	**Female**	$p(MWU)$*	**Male**	**Female**	$p(MWU)$*	**Male**	**Female**	$p(MWU)$*
Android	3	2	0.11	449	338	0.29	1.41	0.52	0.006
Chromium OS	3.85	3.45	0.84	448	663	0.66	2.3	1.75	0.19
Couchbase	10.4	7.5	0.14	1,257	992	0.71	6.4	2.4	0.09
Go	2	6.7	0.06	225	503	0.17	1	2.42	0.15
LibreOffice	3.7	1.83	0.08	517	178	0.06	0.5	0.13	0.09
OmapZoom	3.61	4.38	0.88	329	359	0.85	4.54	4.11	0.62
oVirt	6.85	6.72	0.93	628	358	0.21	6.33	4.18	0.37
Qt	2.86	4.11	0.69	351	718	0.65	3	2.74	0.49
Typo3	1.61	3.66	0.25	230	890	0.28	5.7	2.5	0.31
Whamcloud	1.63	2.20	0.79	370	311	0.78	1.4	3.16	0.83

*Adjusted p-value after applying FDR correction.**

C. Productivity comparison

Based on the two code contribution metrics (i.e., code churn per month and code changes per month), none of the ten projects (Table V) suggest significant differences in productivity between males and females. Figure 5 shows the distributions of code churn per month based on gender for the ten projects. Although none of the differences are statistically significant, Table V suggests twice median code churn per month from females than males in Go, Qt, and Typo3. Based on the results of RQ2, all those three projects have higher ratios of female core developers than the ratios of non-casual female developers (cf. Table IV). On the contrary, an opposite picture can be seen in the LibreOffice project. Unsurprisingly, Table IV also suggests LibreOffice as one of the least gender inclusive projects. Therefore, we believe that as there is not significant difference between male and female productivity, the gender diversity in the projects can promote team productivity and make impact on the ultimate growth in OSS projects.

The only significant difference for this RQ was observed based on code review participation in the Android project, where males were more frequently reviewing other developers' code than females. Since code review participations are primarily based on invitations (i.e., an author selects reviewers for his/her code) [12], females may not be responsible for their lower code review participation, since many of their male colleagues may prefer another male to review their code over a female.

Finding 3: *None of the projects suggest significant differences between male and female developers in terms of productivity as estimated using two code contribution metrics. In an inclusive OSS project, female developers would on average as productive as their male colleagues, with some female developers being more productive than the average male developer.*

D. Explicit biases for / against women

We estimated explicit technical biases for / against females using four observable metrics as defined in Table III. The results of our statistical tests (Table VI) suggest significant explicit biases against females in three projects (i.e., Android, Chromium OS, and LibreOffice). Female developers from those three projects: i) had significantly lower code acceptance rates than males, ii) wait significantly longer than males to get initial feedback for their code changes, and iii) wait significantly longer than males to complete code reviews. Code submitted by females also encounter significantly higher

TABLE VI
EXPLICIT BIASES FOR / AGAINST WOMEN

Project	Acceptance rate			First feedback interval (hrs)			Review interval (hrs)			Churn per comment		
	Male	**Female**	$p(\chi^2)$*	**Male**	**Female**	$p(MWU)$*	**Male**	**Female**	$p(MWU)$*	**Male**	**Female**	$p(MWU)$*
Android	88.1%	77.6%	<**0.001**	1.64	2.25	<**0.001**	20.1	69.5	<**0.001**	9.66	6	<**0.001**
Chromium OS	91.4%	90.1%	**0.001**	1.44	3.2	<**0.001**	31.1	31.3	0.19	6	5	<**0.001**
Couchbase	92.1%	90.4%	0.015	3.75	3.25	0.08	17.5	21.6	<**0.001**	10	15	<**0.001**
Go	91.6%	90.8%	0.44	1.4	1.2	0.1	16.6	16.7	0.46	7.2	7	0.46
LibreOffice	92.3%	83.5%	<**0.001**	14.7	25.5	<**0.001**	23.3	72.4	0.41	10	10.7	0.41
OmapZoom	75.05%	75.8%	0.45	11.5	12.6	0.07	160.1	146.9	0.70	3.66	4	0.08
oVirt	88.1%	91.1%	<**0.001**	4.4	4.5	0.07	49.4	47.9	0.15	3	3	0.99
Qt	90.5%	93.8%	<**0.001**	2.57	1.2	<**0.001**	32.7	19.8	<**0.001**	7	8	<**0.001**
Typo3	91.1%	92.0%	0.36	0.52	0.18	<**0.001**	25.9	12.1	<**0.001**	4.2	6.3	<**0.001**
Whamcloud	78.8%	78.4%	0.89	6.43	4.8	0.07	367.9	356.9	0.54	3.3	2.9	0.063

*Adjusted p-value after applying FDR correction.**

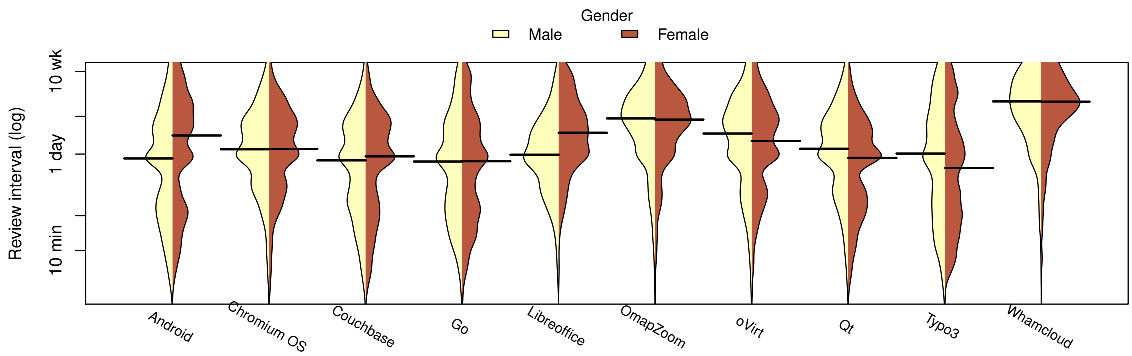

Fig. 6. Review Interval (male vs. female)

scrutiny (i.e., lower churn per comment) in both Android and Chromium OS. The differences between males and females were most prominent in Android and LibreOffice, where women had 10% lower acceptance rates as well as more than three times longer review intervals (Figure 6) than males.

On the contrary, three of the projects (oVirt, Qt, and Typo3) indicate signs of biases favoring female developers. Those include: i) code changes from females were more likely to be accepted than males in oVirt and Qt, ii) females were waiting significantly shorter than males to get initial feedback or complete code reviews for their code changes in Typo3 and Qt, and iii) females' code changes were less scrutinized than those from males in Qt and Typo3. The results are mixed for Couchbase where although females encountered significantly longer review intervals, their code changes were less scrutinized.

Finding 4: *Explicit technical biases exist in seven out of the ten projects in our study. However, the type and direction of biases vary across those projects. Biases against females were observed in three projects while three other projects indicate biases favoring females.*

VI. DISCUSSION AND IMPLICATIONS

In this section, we compare our results with the results of prior studies and discuss the possible reasons and implications of our findings.

A. Characteristics of the biases

The results of our study hint both explicit and implicit biases among OSS projects. Explicit biases such as lower code acceptance, longer first feedback interval, longer review interval, and higher code scrutiny for code are immediately identifiable from the results of RQ3 (Section V-D). While we did not design a metric to identify implicit biases, the results of RQ2 (Table V) hint prejudices among some developers against inviting females to review their codes. In nine out of the ten projects, females participated in less code reviews per month than males; although in four out of those nine projects females committed more code changes per month than males. We hypothesize that the implicit biases among male developers doubting the coding abilities of females may be a factor here.

On the contrary, our results also show that biases against females are not norms in most of the OSS projects. Projects with female-unfriendly cultures do exist but majority of the projects do not foster such a culture. Only three out of the ten projects from our study were biased against females. More promisingly, three other projects were more supportive to females than males. Although, we have a long way to go before establishing OSS projects with equal ratio of female developers, the current picture is not completely bleak. Some of the existing projects are already fostering cultures to promote gender diversity. OSS projects that are actively promoting diversity and inclusion include: Mozilla[5], Debian[6], and Fedora[7].

B. Impacts of biases

In a recent article, one female developer commented, "In regards to the open source projects - I've been thinking about this recently. I actually haven't committed to any and it definitely puts a kink in my career I feel like it's a circle I can't get into. But mostly I fear the excessive spotlight of being a sole female programmer on a publicly available project. In light of how women are treated on the internet, this fear does not seem unreasonable [45]." The impact of conscious or unconscious bias against women may have large impact on women retention rate in OSS projects.

To investigate the impacts of the identified biases in our study, we computed ratios of active female contributors during each six months interval for the last six years for six projects (Android, Chromium OS, LibreOffice, oVirt, Qt and Typo3) that indicated biases for / against females. To better understand the trends, we plotted simple moving average over two year periods (Figure 7). We observed increasing trend for the percentage of female contributors in Chromium OS, Typo3, and Qt. We observed consistently decreasing trend for Android project. LibreOffice showed mixed trend as ratio of female increased during the first three years but declined for the last two years. Since we observed most prominent biases against females in Android and LibreOffice, we hypothesize that these declining trends may be influenced by those biases. In a followup study, we aim to investigate the differences in culture between these two groups of projects (i.e., those that favor women vs. those that have biases against women), to pinpoint the factors that negatively impact women's participation in OSS projects.

C. Importance of female leaders

Lack of female role models has been identified as a major cause behind lower interests among females to join OSS projects, as Sheryl Sandberg said "There aren't more women in tech because there aren't more women in tech. [28]." Our results also suggest that projects that have better ratios of females as core developers (e.g., Qt, Typo3) were not biased against female developers. Moreover, females from those two projects also had higher median code churn per month than males.

Therefore, initiatives to promote females in leadership position may help improving gender diversity among OSS projects. For example, OSS projects can allocate quotas to females for

[5]https://blog.mozilla.org/inclusion/2018/04/19/diversity-and-inclusion-at-mozilla/
[6]https://www.debian.org/intro/diversity
[7]https://docs.fedoraproject.org/en-US/diversity-inclusion/

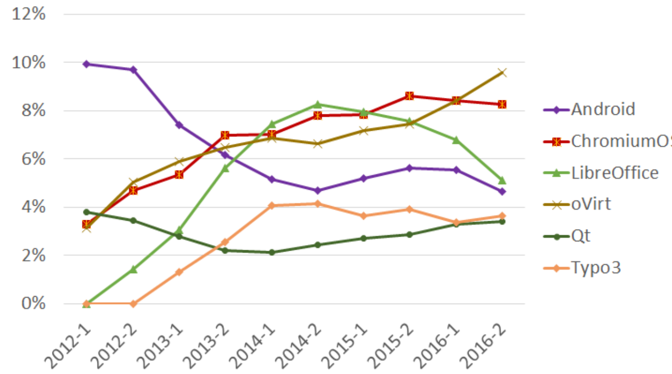

Fig. 7. Percentage of active contributors that are females (moving average over two years)

leadership positions to ensure adequate female role models. OSS projects can also develop special workshops or training programs using female leaders so that they can share their experience and inspire other females to promote themselves to leadership positions. Finally, OSS projects can also highlight the achievements of its women contributors to motivate more women to join the project.

D. Comparison with Terrell et al.'s study

Although Terrell et al.'s study is a significant motivation behind this study, some of the results of our study does not concur with their findings. First, while Terrell et al. found women having significantly higher acceptance rates than males in Github, we observed the same in only two out of the ten projects and three projects had contradictory results. Second, they found women's code changes having larger code churns than females, we found three (i.e., Couchbase, LibreOffice, and oVirt) out of the ten projects supporting that result, two projects (i.e., Android, and Go) contradicting that result, and the remaining five projects showing no significant differences.

While our results also suggest that gender biases do exist in some OSS projects, those biases may not be universal. The prevalence of gender bias may depend on several factors (e.g., project sponsor, community culture, and community commitment to promote diversity). Majority of the OSS projects in our study do not indicate biases against females.

E. Implications and Possible Solutions

Today's world rides on open source software. The web, smart phones, the Amazon Echo, your car - every high-tech depends on open source these days [20]. "I thought open source was the great democratizing force in the world, but it does seem that there is a pervasive pattern of gender bias in tech, and it's even worse in the open source culture." said by Larissa Shapiro, Head of Global Diversity and Inclusion at Mozilla [30]. So if we cannot reduce the bias in open source fields, it will have negative consequences in all other fields of computer science. Although some OSS projects such as Mozilla, Debian, and Fedora have already adopted some measures to improve diversity and inclusion [30], we should focus more on adopting measures that will ensure that women are not underestimated even if their identity is revealed.

In automotive industries, there are some specific groups comprising the women in this field. They regularly publish articles, e-news recognizing leading women personalities in their fields. They have built a platform where women can share their experiences, exchange views, discuss potential solutions of their problems [43]. If we can build similar social network for women in OSS projects, they can share their views and experiences which will ultimately increase their confidence level and self-efficacy in their respective professions. They will be motivated to raise their voices instead of leaving their jobs silently.

The contemporary gender diverse IT industries in some countries such as the Philippines, Brazil, Malaysia and Vietnam, give hope to build a more gender diverse OSS ecosystem. In a recent study [31], Ulf et al. claimed that women dominate computer science in Malaysia due to the intensive ethnification of the Malaysian society and its consequences for higher education. They claimed that the complex interrelationships of gender, class, age and ethnicity in a multi-ethnic society must be investigated in order to construct a women-friendly technological field. In western countries, computer science is numerically dominated by men, but in Malaysia the polarity of the 'charge' has been changed as they followed a strategic plan from the very beginning of their education system through the professional level to reduce gender gap. Therefore, it is crucial to investigate relational aspects of gender, including positional relations of women and men with other relevant social categories intersecting and shaping gender relations, to ensure best practices in technological world.

VII. Threats

In the following subsections, we address three common types of threats to any empirical study.

A. Internal Validity

The primary threat to internal validity in this study is *project selection*. We included ten publicly accessible OSS projects that practice tool-based code reviews supported by the same tool (i.e., Gerrit). Though, it is possible that projects supported by other code review tools (e.g., ReviewBoard, Github pull-based reviews, and Phabricator) could have behaved differently, we think this threat is minimal for four reasons: 1) all code review tools support the same basic purpose, i.e. detecting defects and improving the code, 2) the basic workflow (i.e. authors posting code, reviewers commenting about code snippets, and code requiring approval from reviewer before integration) of most of the code review tools are similar, 3) we did not use any Gerrit-specific feature/attributes in this study, and 4) sentiments expressed in review comments may not depend on any feature that is exclusive to Gerrit only. Therefore, we believe the project selection threat is minimal.

B. Construct Validity

The primary threat to construct validity is related to our gender resolution methodology based on the genderComputer tool, which has been used in several prior SE studies [47], [44]. Since we manually validated the tool's classification for the 937 non-males, we believe that no male was misclassified as a female in our dataset. However, as we accepted genderComputer's classifications for the 2,633 males, we are unable to make a similar claim that no female was misclassified as a male. To estimate possible classification errors, we manually validated the genders of 200 randomly selected developers from the 2,633 males, using a similar methodology that we used for the non-males. Since our validation found only one female (0.5%) among these 200 developers, we do not think enough females were misclassified as males in our dataset to alter the results of our research questions. We also did not attempt to identify non-binary genders, since it is almost impossible to identify those developers without their inputs.

C. External Validity

Although we analyzed a large number of code review requests from ten popular and matured OSS projects, we cannot definitively establish that our sample is representative of the entire OSS population. Because OSS projects vary on characteristics like product, participant type, community structure, and governance, we cannot draw general conclusions about all OSS projects from this single study. To build reliable empirical knowledge, we need family of experiments [8] that include OSS projects of all types.

VIII. Conclusion

In this study, we explored level of gender diversity among ten popular OSS projects. We also explored the existence of gender biases among those projects. Our results suggest that lack of gender diversity remains an ongoing issue among all the ten projects as each of the projects had less than 10% female developers. However, many of the projects also suffer from lack of inclusion of females to leadership positions. Although none of the projects suggest significant differences between male and female developers in terms of productivity based on three different measures, data from three out of the ten projects indicate explicit biases against female developers with lower code acceptance rates as well as delayed feedback during code reviews. However, biases against females are not universal as majority (i.e., seven out of ten) of the projects do not discriminate against females. The two projects with the least ratios of female contributors as core developers showed the most biases against females. Unsurprisingly, those two projects also had declining ratios of female contributors over the last three years. Based on our findings, we conclude that promoting and mentoring females to leadership positions may be an effective solution to promote gender diversity.

In a follow-up study, we intend to investigate the differences in various attributes such as community culture, management structure, leadership composition, project sponsor, and developer recruitment strategy between the projects that show biases against women and the projects that favor women to pinpoint solutions to promote gender diversity.

References

[1] "Floss." [Online]. Available: http://geekfeminism.wikia.com/wiki/FLOSS#Discussion_of_issues

[2] (2017) Girls who code. [Online]. Available: https://girlswhocode.com/

[3] *Oxford English Dictionary*. Oxford University Press, 2017.

[4] B. Alba, "To achieve gender equality, we must first tackle our unconscious biases," 2018. [Online]. Available: http://theconversation.com/to-achieve-gender-equality-we-must-first-tackle-our-unconscious-biases-92848

[5] M. Asay, "Who really contributes to open source," 2018. [Online]. Available: https://www.infoworld.com/article/3253948/open-source-tools/who-really-contributes-to-open-source.html

[6] C. Ashcraft, B. McLain, and E. Eger, "Women in tech: The facts," *National Center for Women et Information Technology*, 2016.

[7] A. Bacchelli and C. Bird, "Expectations, outcomes, and challenges of modern code review," in *Proceedings of the 2013 International Conference on Software Engineering*. IEEE Press, 2013, pp. 712–721.

[8] V. R. Basili, F. Shull, and F. Lanubile, "Building knowledge through families of experiments," *IEEE Transactions on Software Engineering*, vol. 25, no. 4, pp. 456–473, 1999.

[9] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[10] C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan, "Mining email social networks," in *Proceedings of the 2006 international workshop on Mining software repositories*. ACM, 2006, pp. 137–143.

[11] Bitergia, "Gender-diversity analysis of the linux kernel technical contributions," 2016. [Online]. Available: https://blog.bitergia.com/2016/10/11/gender-diversity-analysis-of-the-linux-kernel-technical-contributions/

[12] A. Bosu, J. C. Carver, C. Bird, J. Orbeck, and C. Chockley, "Process Aspects and Social Dynamics of Contemporary Code Review: Insights from Open Source Development and Industrial Practice at Microsoft," *IEEE Transactions on Software Engineering*, vol. 1, no. 99, pp. 1–1, 2017.

[13] A. Bosu and J. C. Carver, "Impact of peer code review on peer impression formation: A survey," in *2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2013, pp. 133–142.

[14] ——, "Impact of Developer Reputation on Code Review Outcomes in OSS Projects: An Empirical Investigation," in *Proceedings of the 2014 ACM / IEEE International Symposium on Empirical Software Engineering and Measurement*, 2014, pp. 33:1–33:10.

[15] K. Campbell and A. Mínguez-Vera, "Gender diversity in the boardroom and firm financial performance," *Journal of business ethics*, vol. 83, no. 3, pp. 435–451, 2008.

[16] K. Crowston, K. Wei, Q. Li, and J. Howison, "Core and periphery in free/libre and open source software team communications," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, vol. 6. IEEE, 2006, pp. 118a–118a.

[17] J. Damore. Google's ideological echo chamber: How bias clouds our thinking about diversity and inclusion. [Online]. Available: https://gizmodo.com/exclusive-heres-the-full-10-page-anti-diversity-screed-1797564320

[18] L. DiDio, "Look out for techno-hazing," vol. 31, no. 39, pp. 72–73.

[19] N. L. Ensmenger, *The computer boys take over: Computers, programmers, and the politics of technical expertise*. Mit Press, 2012.

[20] K. Finley, "Diversity in open source is even worse than in tech overall," 2017. [Online]. Available: https://www.wired.com/2017/06/diversity-open-source-even-worse-tech-overall/

[21] S. Fowler. (2017) Reflecting on one very, very strange year at uber. [Online]. Available: https://www.susanjfowler.com/blog/2017/2/19/reflecting-on-one-very-strange-year-at-uber

[22] R. A. Ghosh, R. Glott, B. Krieger, and G. Robles, "Free/libre and open source software: Survey and study," 2002.

[23] C. Herring, "Does diversity pay?: Race, gender, and the business case for diversity," *American Sociological Review*, vol. 74, no. 2, pp. 208–224, 2009.

[24] S. A. Hewlett, L. Sherbin, F. Dieudonne, C. Fargnoli, and C. Fredman, "Athena factor 2.0: Accelerating female talent in science, engineering & technology," *Center for Talent Innovation2014*, 2014.

[25] T. James, M. Galster, K. Blincoe, and G. Miller, "What is the perception of female and male software professionals on performance, team dynamics and job satisfaction?: insights from the trenches," in *Proceedings of the 39th International Conference on Software Engineering: Software Engineering in Practice Track*, 2017, pp. 13–22.

[26] M. Joblin, S. Apel, C. Hunsen, and W. Mauerer, "Classifying developers into core and peripheral: An empirical study on count and network metrics," in *2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)*. IEEE, 2017, pp. 164–174.

[27] V. Kuechler, C. Gilbertson, and C. Jensen, "Gender differences in early free and open source software joining process," in *IFIP International Conference on Open Source Systems*. Springer, 2012, pp. 78–93.

[28] J. Lansing, "Women in tech: don't even try to fit in a man's world," 2015, accessed = 2018-05-19. [Online]. Available: http://fortune.com/2015/08/08/jane-lansing-women-in-tech/

[29] S. Levin. (2017) Female engineer sues tesla, describing a culture of 'pervasive harassment'. [Online]. Available: https://www.theguardian.com/technology/2017/feb/28/tesla-female-engineer-lawsuit-harassment-discrimination

[30] J. McConnell, "Mozilla experiment aims to reduce bias in code reviews," 2018. [Online]. Available: https://blog.mozilla.org/blog/2018/03/08/gender-bias-code-reviews/

[31] U. Mellstrm, "The intersection of gender, race and cultural boundaries, or why is computer science in malaysia dominated by women?" *Social Studies of Science*, vol. 39, no. 6, pp. 885–907, 2009.

[32] A. Mockus, R. T. Fielding, and J. D. Herbsleb, "Two case studies of open source software development: Apache and mozilla," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 11, no. 3, pp. 309–346, 2002.

[33] D. Nafus, J. Leach, and B. Krieger, "Gender: Integrated report of findings," *FLOSSPOLS, Deliverable D*, vol. 16, 2006.

[34] u. . h.-s.-w.-r. Nathan Ingraham. (2017) Report: Apple is a sexist, toxic work environment.

[35] M. Ortu, G. Destefanis, S. Counsell, S. Swift, M. Marchesi, and R. Tonelli, "How diverse is your team? investigating gender and nationality diversity in github teams," *PeerJ Preprints*, vol. 4.

[36] A. Panteli, J. Stack, M. Atkinson, and H. Ramsay, "The status of women in the uk it industry: an empirical study," *European Journal of Information Systems*, vol. 8, no. 3, pp. 170–182, 1999.

[37] I. Papuc, "Is open source open to women?" 2017. [Online]. Available: https://wotc.crn.com/blog/is-open-source-open-to-women

[38] R. Paul, A. Bosu, and K. Z. Sultana, "Expressions of sentiments during code reviews: Male vs. female," in *Proceedings of the 26th IEEE International Conference on Software Analysis, Evolution and Reengineering*, ser. SANER '19, 2019.

[39] G. Robles, J. M. Gonzalez-Barahona, and I. Herraiz, "Evolution of the core team of developers in libre software projects," in *2009 6th IEEE international working conference on mining software repositories*. IEEE, 2009, pp. 167–170.

[40] G. Robles, L. A. Reina, J. M. González-Barahona, and S. D. Domínguez, *Women in Free/Libre/Open Source Software: The Situation in the 2010s*, 2016.

[41] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[42] F. Sharan, "Asf committer diversity survey - 2016," 2017. [Online]. Available: https://cwiki.apache.org/confluence/display/COMDEV/ASF+Committer+Diversity+Survey+-+2016

[43] M. Sultana, S. Shafii, and S. Hussain, "Women professionals in automotive industry through international college of automotive (icam), pekan, malaysia," *IOSR Journal of Business and Management*, vol. 14, no. 1, pp. 50–55, 2013.

[44] J. Terrell, A. Kofink, J. Middleton, C. Rainear, E. Murphy-Hill, C. Parnin, and J. Stallings, "Gender differences and bias in open source: Pull request acceptance of women versus men," *PeerJ Computer Science*, vol. 3, p. e111, 2017.

[45] N. Tiku, "This is why there aren't enough women in tech," 2013. [Online]. Available: http://valleywag.gawker.com/this-is-why-there-arent-enough-women-in-tech-1221929631

[46] B. Vasilescu, A. Capiluppi, and A. Serebrenik, "Gender, representation and online participation: A quantitative study," *Interacting with Computers*, vol. 26, no. 5, pp. 488–511, 2014.

[47] B. Vasilescu, D. Posnett, B. Ray, M. G. van den Brand, A. Serebrenik, P. Devanbu, and V. Filkov, "Gender and tenure diversity in github teams,"

in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 3789–3798.

[48] B. Vasilescu, A. Serebrenik, and V. Filkov, “A data set for social diversity studies of github teams,” in *Proceedings of the 12th Working Conference on Mining Software Repositories*. IEEE Press, 2015, pp. 514–517.

[49] Y. Ye and K. Kishida, “Toward an understanding of the motivation open source software developers,” in *Proceedings of the 25th international conference on software engineering*. IEEE Computer Society, 2003, pp. 419–429.

[50] F. Zlotnick, “Github open source survey 2017,” http://opensourcesurvey.org/2017/, Jun. 2017.