

Gender Differences in Early Free and Open Source Software Joining Process

Victor Kuechler, Claire Gilbertson, Carlos Jensen

► **To cite this version:**

Victor Kuechler, Claire Gilbertson, Carlos Jensen. Gender Differences in Early Free and Open Source Software Joining Process. 8th International Conference on Open Source Systems (OSS), Sep 2012, Hammamet, Tunisia. pp.78-93, 10.1007/978-3-642-33442-9_6 . hal-01519076

HAL Id: hal-01519076

<https://hal.inria.fr/hal-01519076>

Submitted on 5 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Gender Differences in Early Free and Open Source Software Joining Process

Victor Kuechler, Claire Gilbertson, Carlos Jensen

School of Electrical Engineering and Computer Science
Oregon State University, Corvallis, OR, USA
{kuechlej@onid.orst.edu, claire.gilbertson@gmail.com,
jenseca@eecs.orst.edu}

Abstract. With the growth of free and open source software (FOSS) and the adoption of FOSS solutions in business and everyday life, it is important that projects serve their growingly diverse user base. The sustainability of FOSS projects relies on a constant influx of new contributors. Several large demographic surveys found that FOSS communities are very homogenous, dominated by young men, similar to the bias existing in the rest of the IT workforce. Building on previous research, we examine mailing list subscriptions and posting statistics of female FOSS participants. New participants often experience their first interaction on a FOSS project’s mailing list. We explored six FOSS projects – Buildroot, Busybox, Jaws, Parrot, uClibc, and Yum. We found a declining rate of female participation from the 8.27% of subscribers, to 6.63% of posters, and finally the often reported code contributor rate of 1.5%. We found a disproportionate attrition rate among women along every step of the FOSS joining process.

Keywords: joining process, diversity, mailing lists

1 Introduction

Although a similar percentage of men and women receive bachelor’s degrees today [19], there is a significant difference in the percentage for computer science and engineering. Only 25% of IT workers are women [20], and women earn around 18% of IT-related bachelor’s degrees [28]. A smaller percentage of women actively participate in FOSS, less than 2% [7, 9, 14, 18, 24]. Why do so few women participate in FOSS, and what can attribute to these differences?

FOSS projects need to attract and maintain active users. The volunteer nature of FOSS and the general lack of financial incentives to participate lead to high turnover, and the need for a continuous influx of developers [29]. The transition from user to contributor begins on a mailing list. Mailing lists are at the heart of all communication and discussion in FOSS projects, and therefore at the heart of all FOSS projects. Its archival nature also preserves past, present and future design and implementation decisions, as well as the project’s evolving culture.

Joining a FOSS project often begins with lurking or silently observing the community by subscribing to its mailing list. Since the majority of communication occurs on mailing lists, we need to understand how this first step affects newcomers’ motivations and future behavior.

A previous study found that almost 80% of newbies received a positive reply to their first post, and those who received a timely response were more likely to contin-

ue participating [11]. This study also found that messages from men and women were treated similarly in terms of tone, helpfulness, and likelihood of replies, yet significantly fewer women posted (2.68%). The study did not address the time users lurked on mailing lists before posting, which may be an influential factor.

Building on previous research, we examined subscriber logs and data for six FOSS projects: Buildroot, Busybox, Jaws, Parrot, uClibc, and Yum. We examined the differences between posters and non-posters to determine the attrition rate of women at different stages of the joining process. More specifically we sought to examine the following research questions:

- RQ1: Once subscribed to a FOSS mailing list, are women as likely to participate (post) as men?
- RQ2: Do women participate (post) with the same frequency as males?
- RQ3: Do women lurk longer than men before posting?
- RQ4: Do men and women participate (subscribe) for equal amounts of time?

The paper is structured as follows: First we review work related to FOSS communities and project joining. Next, we describe our methodology for collecting and analyzing data. In section 4 we describe our results. In section 5 we discuss our research questions. We finish by reviewing our data and presenting our conclusion.

2 Related Work

The influence of FOSS has grown over the last decades and shows that FOSS software can be more reliable and perform better than proprietary software [27]. FOSS encompasses a great variety in projects, from the highly technical Linux Kernel, supporting operating systems like Ubuntu, Fedora and Debian, to end-user applications like Android, Wikipedia, and business solutions like Open Office and the GIMP.

FOSS is a volunteer-driven development paradigm that brings together developers and contributors from around the world. Only 30% of developers are paid [7, 14] and what motivates contributors can be both intrinsic and extrinsic [5, 14]; a majority of FOSS developer surveys find that contributors are motivated by the opportunity to improve their programming skills [7, 14, 24]. Working on FOSS allows users of any age, education, or experience level to gain valuable skills. One study shows that a significant number of people “wished to improve software products for other developers,” [9] and another finds that 77% of respondents thought giving back to the community is very important [14]. A passionate user may also start their own project because he or she has an unaddressed interest or need that could be met by a FOSS solution [22]. FOSS projects are also expanding to address humanitarian needs, which attract a different kind of developer [1].

Surveys of FOSS developer demographics, although outdated, show that FOSS communities are 98% male [7, 9, 14, 18, 24] with an average age of 27 [7, 9, 14, 24]. On average, FOSS developers are highly educated with 30% having at least a bachelor’s degree and 10.6% a graduate degree. Most contributions are from Western Eu-

rope and the United States. Many FOSS communities have taken steps to address inequalities by starting different programs aimed at recruiting and retaining underrepresented groups. While some projects acknowledge the importance of different types of diversity, most focus on recruiting and retaining female contributors.

There are several text-based forms of communication used in FOSS projects to maintain project awareness. Internet Relay Chat (IRC) is used for real-time communication. Asynchronous communication in the form of blogs, wikis, forums, and bug-tracking systems archive project discussions, documentation and project news. These channels tend to augment mailing lists, the primary medium used for interaction in FOSS communities [10]. This tool allows everyone to participate asynchronously, keep up-to-date with new developments, bounce ideas back and forth, and encourage discussions about the project. Gutwin et al. note that “there is a strong culture of ‘making it public’ [in FOSS] where developers are willing to answer questions, discuss their plans, report on their actions, and argue design details, all on the mailing list” [10].

Several studies focus on mailing list activity and its influence on future participation. Lampe and Johnston examine the Slashdot community over a one-month period and found that more than 55% of newcomers made only one comment and those that received no feedback were less likely to continue to comment [15]. Krogh et al. studied Freenet’s developer list for one year and find that of successful joiners, over three-quarters started a new thread, and that the 10% of participants who never received replies dropped out [13].

Jensen et al. focus on new users instead of developers [11]. They observe the first posts made by new users and how existing members greeted them. After examining the newbie’s tone, nationality, and gender they find that newbies received equally prompt replies, but flaming or aggressive replies were not uncommon. They also argue that because this behavior is public it could have far-reaching effects: “Thus, while OSS participants were generally polite to newbies, it is possible that newbie expectations and perceptions of politeness could be colored by how the regulars engage with each other” [11]. In other words, lurkers may be pushed away based solely on observed negative behavior on mailing lists.

Mailing lists keep an exact record of public discussions. “Mailing lists allow people to find out who the experts are, simply by initiating a discussion: because the messages go to the entire group, the ‘right people’ will identify themselves by joining the conversation.” [10]. This makes mailing lists a great source of information, and subscription logs let us see not only who uses the list, but also who is lurking to test the waters.

FOSS communities are hierarchical in nature and the Onion Model in Figure 1, developed by Ye and Kishida [29], shows a very simplistic hierarchy of roles in FOSS communities. Newcomers often begin at the outer layers as passive users who may have subscribed to a mailing list or IRC channel. It is difficult to understand lurker behavior, in part due to the difficulty determining who is a lurker.

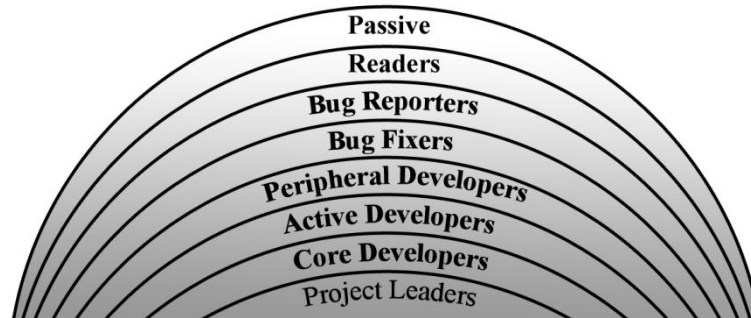


Fig. 1. Onion Model of FLOSS joining process [29]

Nonnecke and Preece examined lurking as a transitional phase needed in order for users to feel comfortable contributing to a technical discussion. They find that the transition could last anywhere from weeks to months, but they did not examine the effects of gender on this process [21]. Other studies, mostly of discussion forums and mailing lists, show lurkers make up from 50% [21, 25] to 90% [12, 16] of an online community. Lurking is recommended to newcomers in FOSS communities as a method to learn the current state of the project, who to talk to, what channels of communication to use, and project culture.

3 Methodology

3.1 Data Collection

The Oregon State University Open Source Lab (OSL) hosts over 150 projects and “distribute software to millions of users globally” [2]. The OSL acted as agents in gaining access to mailing list data and user subscription logs, with projects’ consent. We selected active projects with many users and high traffic.

We also asked numerous other communities (not hosted at the OSL) for access to their data, with mixed results; some communities were willing to share their data but did neither use mailing lists nor keep subscription logs. Some were wary about sharing user’s information, seeing potential privacy issues. Without a complete set of message files and corresponding logs, we could not use these projects for comparison.

The data we collected was all from mature, highly technical projects. These data sets spanned at least 500 days and have between 73 and 944 subscribers per mailing list.

3.1.1 Project Descriptions

We used data from six FOSS projects: Buildroot, Busybox, Jaws, uClibc, Yum and Parrot. From these we selected eleven mailing lists: Buildroot, Busybox, Jaws,

Jaws-announces, Jaws-bugs, Jaws-commits, Jaws-developers, Jaws-general, uClibc, Yum, Yum-devel, and Parrot. Table 1 shows the time coverage of each mailing list.

Buildroot is a technical project that helps users install Linux on embedded systems [6]. Buildroot has one mailing list and a very active community that communicates about commits, questions, bug reports, and patches.

Busybox is a highly technical project that merges numerous UNIX utilities commonly found in GNU fileutils, shellutils and others [2], specifically for embedded systems [3]. This project has one mailing list, busybox, which is the main source of communication and the suggested medium for communicating with the community.

Jaws is a technical framework and content management system (CMS) that encourages users to develop their own modules. We included this project because it was smaller than others and appeared to cater to a broader user base (anyone with web authoring skills). This project contained multiple mailing lists, which we combined to more effectively compare with the others. We analyzed at Jaws, Jaws-announces, Jaws-bugs, Jaws-commits, Jaws-developers, and Jaws-general.

Parrot is designed to compile byte code for dynamic languages. Parrot’s website directs users to the parrot-dev mailing list for development and discussion. Other documentation sends new users to parrot-users, which is practically unused [26].

uClibc or the “microcontroller C library” [4] is a smaller alternative to the GNU C Library, and almost all applications supported by glibc are compatible. uClibc has two mailing lists—one for discussion and development (uClibc) and another for source commits (uClibc-cvs), which was dedicated to different files for bug patches and other code changes. We chose to examine the list for discussion and development since this is where new users are most likely to interact with the community.

Yum or “yellowdog updater modified” is a package management system that provided tools to automate software installation, upgrading, configuring and uninstallation [30]. Yum works with RPM-based Linux distributions. This project had four mailing lists: rpm-metadata, yum, yum-commits, and yum-devel. We chose to examine yum-devel and yum since these mailing lists were more active and included a variety of users, including newcomers. The commits list mostly included code modifications and updates with few questions or other communication. The rpm-metadata list was not available at the time.

Table 1. Time period of each mailing list subscription log per project

Projects	# of subs	Start	End	Days
Buildroot	944	11/20/08	10/12/10	691
Busybox	695	11/20/08	05/18/10	544
Jaws	73	Announces	11/02/10	1085
		Bugs	08/30/10	1063
		Developers	10/07/10	1059
		General	11/02/07	1086
		Commits	11/03/10	1087
Parrot	698	07/30/08	05/16/10	989
uClibc	428	12/04/08	05/18/10	529
Yum	360	09/26/08	05/13/10	594
Yum-devel	112	09/26/08	05/18/10	599
All lists (mean)	3310	N/A		~600

3.2 Data Parsing

Using documentation from QMAIL [17], we created a Java program to parse the MBOX files and subscription logs to extract the following data (when available):

- Email address
- First name
- Last name
- Subscribe date (if any)
- Unsubscribe date (if any)
- Time on mailing list (Unsubscribe – Subscribe date)
- Number of posts
- Gender
- Date of first post
- Time spent on the list before first post (First post – subscribe date)
- Last Post
- Frequency of posts [(Last - First Post) / (number of Posts)]
- List (used when combining data from multiple lists)

The program iterated over users in the subscription logs for each MBOX file and counted the number of posts made by that person. When the users signed up for the list they could choose to add a first and/or last name in addition to the required email address. We used this information to determine the gender of the subscriber. When available we parsed this information from the logs. When this was not available, we attempted to extract a name from the email address. We used pattern matching to find possible names using the following patterns:

First.last@...
 First_last@...
 First-last@...

If these schemes did not offer a match, we added the entire *username* portion of the email to the *first name* field.

We recorded users' mailing list subscribe and unsubscribe dates; some users subscribed and unsubscribed multiple times. In these cases, we treated the first subscription as the join date and the last un-subscription as the unsubscribe date. Some users did not have a subscription or unsubscribe date in the logs. Since we only had data for approximately two years for each list, some long-time subscribers did not join or leave during that time period. For these we assigned a join and leave date based on their earliest and latest activity. For this study, we focused on newcomer behavior and lurking, so these experienced subscribers were less of an interest. We calculated the total hours spent on the list for each subscriber.

For each user, we counted how many posts, if any, he or she contributed. We did not thread the posts or group them in any way; each post whether a reply or a new topic was counted. For each poster, we recorded the date of their first and last post. From these data points, we determined the amount of time they lurked before posting and their posting frequency.

Using data from the U.S. Census, we matched names to lists of the most common female and male names. We identified 666 users using this process. Some common names are used for both women and men, for instance Alex, Robin, or Morgan. In these cases, we looked at the frequency of use for each gender for each name. If there was a disproportionate use in one gender, we assigned all users with that name to that gender. For example, Alex is ranked as the 63rd most common name for males in the United States, and 990th for females. Therefore, all Alex's were assumed to be male. In cases where the rankings were close, we put the user in an "unknown" category.

Next, we manually filtered obvious "non-names" such as `thepirate@yahoo.com` and identified possible names that did not make it on the list of common U.S. names, or names that did not follow the aforementioned patterns. These names were then shown to other researchers and international students via a web application. These "reviewers" could assign a gender, mark the email address as "not a name," label it as an "unknown" for ambiguous names, or skip the name. We asked reviewers to only assign a sex where they were 100% certain. In the end, we identified 1594 users as either male or female, and were left with 975 unidentified users. Grouping unknown and "not a name" together, 41.66% of subscribers were unidentifiable. While this is unfortunate, we believe this represents a good effort and the users identified were a significant and representative sample of the overall community.

Within this dataset were many extreme values; many users contributed little, and a few users contributed a lot. To normalize the data we arranged the users in each mailing list by the number of posts. If we found a jump of more than an order of magnitude between a user and the next highest contributor, we set this as a cutoff point and excluded the user from our set. We did this to prevent a handful of very frequent posters skewing our statistics. Table 2 shows the number of users excluded from each mailing list.

In addition to treating each list separately, we combined all in order to compare data across FOSS mailing lists. As some of the projects are commonly used jointly

(Busybox, Buildroot, and uClibc) and projects have multiple lists (e.g., Yum and Yum-devel) it is possible that users were counted multiple times in these comparisons.

Table 2. Number of outliers excluded from each data set

Project	Buildroot	Busybox	Jaws	Parrot	uClibc	Yum	Yum-devel	All Lists
Male	0	0	2	0	0	0	0	2
Female	1	0	0	1	1	0	0	3
Unknown	0	0	2	0	0	4	4	10
Total	1	0	4	1	1	4	4	15

4 Results

This section explores our finding and relates them to our research questions. First, we begin by looking at the gender of the subscribers. Secondly, we examine the time subscribers spend lurking before their first post. Next, we present the posting frequency by gender of subscribers and lastly, inspect the amount of time users subscribe to the mailing lists.

4.1 R1: Gender of Subscribers and Posters

Given that women participate at a disproportionately low rate, even by IT standards, can we determine how early in the FOSS joining process these differences emerge? We know from the work of Jensen et al. that by the time a user posts, only 3% of posters are women [11].

In order to answer this question, we counted the number of women and men who subscribed to each mailing list; we found 1769 men and 162 women. 91.73% of all subscribers were male and 8.27% were female. This was more than a 50% decrease when compared to the 20% rate of women in IT, but still much higher than the population of women who contributed code to FOSS projects. Table 3 shows a breakdown of our findings.

Table 3. Number of men, women, and unknown subscribed to each mailing list

Project	Buildroot	Busybox	Jaws	Parrot	uClibc	Yum	Yum-devel
Male	556	423	48	289	218	177	58
Female	52	29	3	27	30	17	4
Unknown	336	243	22	382	180	166	50
Total	944	695	73	698	428	360	112

Are women as likely to participate (post) as men once subscribed? Table 4 contains an overview of the data. The percentage of female posters ranged from 0 to 10.58% of the total number of subscribers, with an average of 6.63%. This was a statistically

significant decrease from the expected value of 8.37% of subscribers ($\chi^2=5.30$, $p=0.0213$). 110, or 67.90% of women never post after joining a mailing list. In comparison, 1065, or 59.30 % of men never posted after they joined a mailing list.

Table 4. Number of men, women, and unknown posters to each mailing list

Project	Buildroot	Busybox	Jaws	Parrot	uClibc	Yum	Yum-devel	All Lists
Male	254	208	9	58	93	80	29	731
Female	21	8	0	4	11	6	2	52
Unknown	157	115	1	47	83	73	23	499
Total	432	331	10	109	187	159	54	1282

4.2 Posting Frequency

Do women post as frequently as men on these lists? In order to determine the number of hours between posts, we examined the time between a user’s first and last post and divided this number by the number of posts for that particular user, see Equation 1. We looked at 563 users who posted at least twice. Table 5 shows the average posting frequency for each list, and for the combined data set.

Table 5. Average posting frequency of men, women, and unknown per mailing list

Lists	Male (hours/post)	Female (hours/post)	Unknown (hours/post)	Δ M-F	Average (hours/post)
Buildroot	289.96	304.29	320.73	-14.32	301.03
Busybox	326.31	246.67	321.74	79.64	322.76
Jaws	168.67	N/A	N/A	N/A	168.67
Parrot	525.33	527.50	490.61	-2.17	512.47
uClibc	393.24	409.63	825.59	-16.39	565.80
Yum	399.26	115.20	749.36	284.06	541.60
Yum-devel	202.10	133.00	722.53	69.10	419.83
All combined	341.70	306.90	495.17	34.81	395.91

$$\frac{1}{n} \left[\sum_{user=1}^{user=n} \frac{(Last Post_{user} - First Post_{user})Hours}{Number of Posts_{user}} \right] \quad (1)$$

Equation. 1. Equation used to calculate the average number of hours between posts

The majority of our lists (excluding Jaws and Yum) showed that men and women posted equally as often. Statistically we did not find any significant difference be-

tween men and women here. We broke separated this set of data into more categories by looking at users who posted at least once, more than once, more than twice, etc.

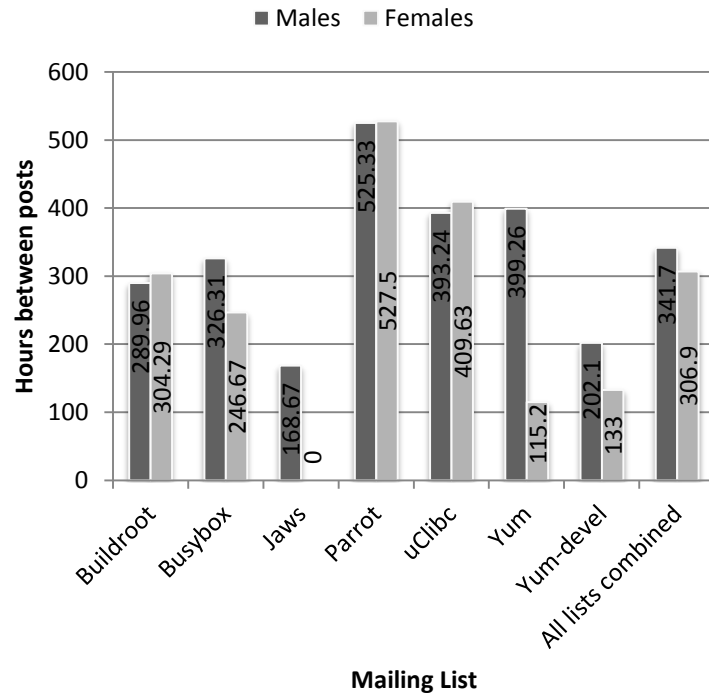


Fig. 2. Posting frequency by hours between posts

We also observed that as we look at the “stickiness” of the community, which is measured by continued participation on the mailing list, the participation of women decreased. Women make up about 6% of posters who submitted between 1 and 3 posts, about 4% of those who submitted between 4 and 8 times, about 2% of those who submitted 9 or 10 times, and 1% submitted more than 10 times. It is interesting to see a near linear trend in the data. Statistical tests show that the proportion of women posters decreased over time ($\chi^2 = 30.346$, $p = 0.0107$). This shows that over time women posted less often, which demonstrates that somewhere along the line women are discouraged from continued participation.

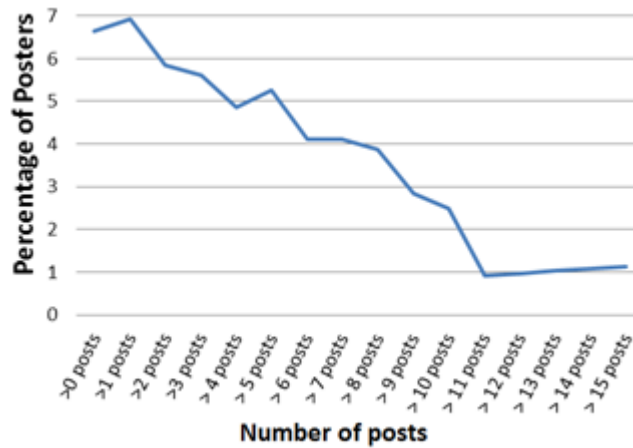


Fig. 3. Posting frequency by hours between posts

4.3 Lurking Habits

The time spent observing a group before contributing is a formative experience that can encourage or discourage users. In most cases, women lurk less than men (note, no female posters were found in the Jaws data set and so it was excluded from the study). The Parrot mailing list turned out to be an extreme outlier; women lurked 2,235.00 hours, which was more than 50% longer than men who lurked 1,406.47 hours. The data from this Parrot (for both men and women) was an extreme outlier, with averages more than 2 standard deviations from that of the other lists. Parrot was therefore excluded from the analysis. A t-test analysis of the two populations (excluding Jaws and Parrot) showed no statistically significant differences in the mean lurking times (p-value=1.72).

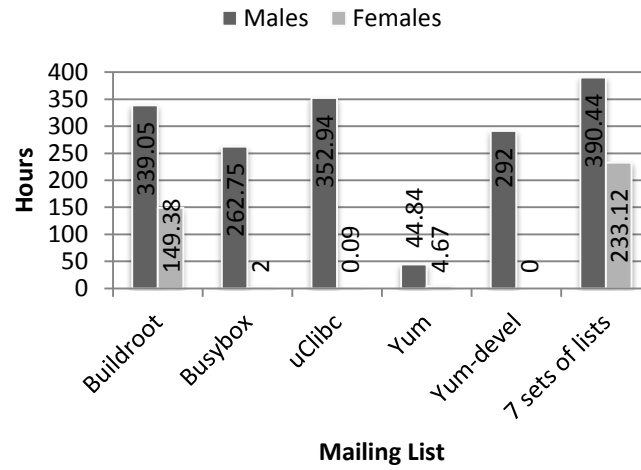


Fig. 4. Posting frequency by hours between posts

4.4 Subscription Length

Do men and women participate (stay subscribed) for similar lengths of time? Interestingly, women and men subscribed for a similar length of time (except in Yum-devel and Parrot), and there were no statistically significant differences.

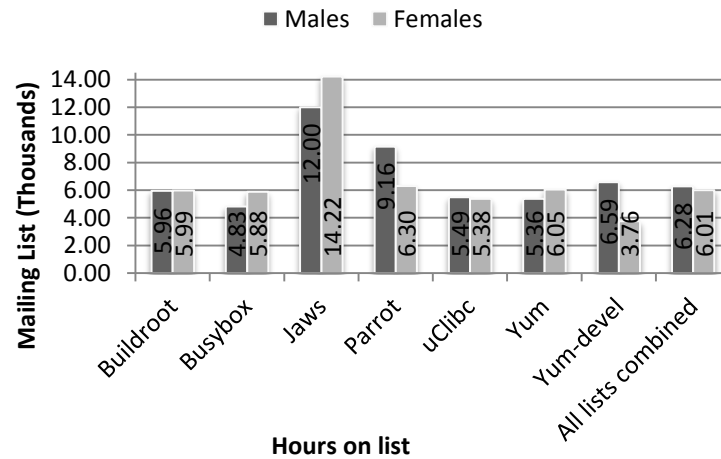


Fig. 5. Subscription length by hours on list

5 Discussion

Jensen et al. [11] found that about 3% of posters in their sample were women. However they focused on a variety of attributes, and gender was not the primary focus. As a consequence, their data had a much larger number of unknown participants, which may have skewed their data. We spent more time manually identifying users as male, female or unknown. We believe our results are in line with other studies that found between 1.5% and 2% of code-contributors are female [7, 9, 14, 18, 24]. We know that the FOSS joining process is complex and the commitment time needed to move into a developer role often excludes newbies with family or social commitments, something that may disproportionately affect women.

Building upon the findings in Jensen et al. [11], we found that 8.39% of the FOSS mailing list subscribers were women. This was significantly lower than the 20% of women in IT. In fact, we discovered an attrition rate throughout the joining and lurking process. Figure 6 shows the decreasing participation of women in FOSS communities.

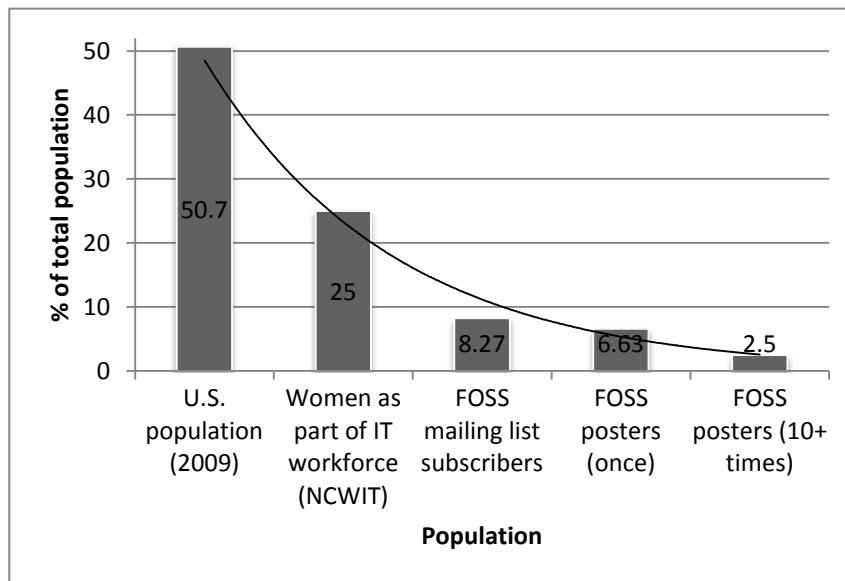


Fig. 6. Women in IT and FOSS

Where and why do gender differences emerge? Joining a mailing list is the first step and location in the FOSS joining process that we can collect data. There are however, other non-documented steps: exploring a project's website, scanning documentation, downloading source code, chatting on IRC, or exploring forums, messages boards or wikis. This leaves a variety of interactions prior to, or in parallel with, the mailing

lists that may also influence women to turn away from FOSS projects. Of those who subscribe to a mailing list, 67.90% of women and 59.30% of men never post, and in our sample only 6.63% of posters were identifiable as women. We found that the number of women sequentially decreases to just over 1% as the number of posts grow to 10, which is in the range of what has been found by other studies.

Over the last ten years, there has been a push to increase the amount of contributor diversity in FOSS. FOSS projects like Dreamwidth have also appeared and managed to attract a 75% female contributor base [23]. However, from our data we cannot see that these efforts are having a strong global effect. More data is needed to determine this however, since the last major FOSS developer survey was performed in 2006 [18].

On average, men posted every 341.70 hours and women 306.90 hours. This shows us that although there is a high attrition rate among women, they are not being excluded from the FOSS conversation. That is not to say there are not many situations in FOSS where women have felt uncomfortable, excluded or specifically targeted in a demeaning way [8]. However, if all women were being forced away, we would expect women to post less frequently than males. What is interesting is that along every step of the joining process, we lose a disproportionate number of women. However, given the small sample of women and our choice of projects, it is difficult to draw statistical conclusions about the causes.

The all subscription logs varied in duration. In particular, Jaws and Parrot covered about 1,000 days, and we found that, on average, users from these two lists subscribed longer than users on other mailing lists, which covered less than 700 days. On nearly all of the lists, women subscribed for slightly less time than males, however we did not find any statistically significant correlation between gender and duration of subscription.

After examining posting statistics, we found that the only statistically significant differences between men and women were the average number of posts and the number of women who kept posting declined more sharply than men. We did not examine the type of messages posted, and it is possible that many users were not interested in joining the project, but rather asked one-time questions. It is unclear why women would be more likely to fall into this category than men. Adding a message-type category to this line of investigation would be beneficial. In addition, the projects we chose are highly technical and therefore comparing them with less technical projects may yield other results.

What drives women away from FOSS in disproportionate numbers? After disproving a number of hypotheses, we are left with two likely factors: women are driven by different sets of motivations and cost-benefit tradeoffs than men or, the social dynamics in projects are more unappealing or hostile to women. As documented by Jensen et al. [11] this may not be blatant or intentional, but the kind of public flaming and aggression documented could be enough to distort participation among a small minority, such as women who may already be hesitant about how they will be received. This does not apply to all women and all projects, since there are many examples of individual success.

Most importantly, what our study shows is that this problem is likely not technical, because most women drop out of FOSS early in their project membership. Efforts to address diversity in FOSS should therefore focus on the first social experiences through programs such as mentorships and making sure novices find the help and support they need.

In the future it would be interesting to evaluate consumer or corporate-oriented projects hosted from a variety of locations. This study used data from very technical projects, which were all hosted in the same location. Also these projects had a similar number of users contributing, and incorporating smaller and larger projects may yield different and possibly more representative results. A system to categorize posts from newbies might add insight about a user's intention to join a FOSS community or otherwise.

6 Limitations

This study used data from very technical projects of similar size, all hosted in the same location (osuosl.org). Our results may have been different if we had sampled less technical, or smaller or larger projects. Lastly, naming conventions, and the subjective nature of matching name and gender could have introduced errors. We did, however, ask reviewers to only assign a gender if they were 100% certain. Studies have shown that women are more likely to try to obscure their gender online, so our analysis may have been skewed.

7 Conclusions

Understanding the reasons behind the gender struggle in FOSS will lead to policies and strategies that encourage greater diversity in FOSS communities. As more companies adopt FOSS software, and FOSS projects diversify to serve a broader population, supporting community infrastructure will be a vital in addressing the issues related to the lack of women in FOSS.

We studied eleven mailing lists and corresponding subscription logs from six FOSS projects, with a combined 3,310 users, of which 1,769 were males, 162 females and 1,379 unknown. We found 8.39% of subscribers were women, which is less than half of the expected 20% population of women in IT. Only 6.63% of posters were women, and the only significant difference we found, in terms of behaviors of men and women, was the average number of posts. Another important finding was the proportion of women who made frequent posts. The percentage of women who posted at least 10 times decreased to about 1%.

On average, males lurked slightly longer (390.44 hours) than females (233.12 hours) before posting to the mailing list for the first time. Also, males subscribed about 270 hours longer than women (less than 5% difference). However, we did not find any statistically significant values in these averages.

Acknowledgements

Thanks are due to the HCI group at Oregon State University, OSU's Open Source Lab, and the project leaders of Buildroot, Busybox, Jaws, Parrot, uClibc, and Yum, for their willingness to participate in this study.

References

1. "About Us." The Humanitarian FOSS Project. <http://www.hfoss.org/index.php/about-us>.
2. "About OSUOSL." OSUOSL. <http://osuosl.org/about-osuosl>.
3. Andersen, E. "BusyBox." <http://busybox.net>.
4. Andersen, E. "UClibc." <http://uclibc.org>.
5. Bitzer, J., Schrettl, W. and Schroder, P. Intrinsic Motivation in Open Source Software Development. *Journal of Comparative Economics*, 35, 1 (2007), 160-69.
6. Buildroot. <http://buildroot.uclibc.org/>.
7. David, P., Waterman A. and Arora, S. "The Free/Libre/Open Source Software Survey for 2003." Stanford University, 2003. <http://www.stanford.edu/group/floss-us/>.
8. Fisher, A. and Margolis, J. "Unlocking the Clubhouse: the Carnegie Mellon Experience." In *Proc. SIGCSE Bulletin* 34 (2002): 79-83.
9. Ghosh, R.A., Glott, R., Krieger, B. and Robles, G. "Free/Libre and Open Source Software: Survey and Study, Part 4: Survey of Developers." June 2002. www.flossproject.org/report/
10. Gutwin, C., Penner, R. and Schneider, K. Group Awareness in Distributed Software Development. In *Proc. CSCS '04*. ACM Press (2004), 72-88.
11. Jensen, C., King, S. and Kuechler, V. "Joining Free/Open Source Software Communities: An Analysis of Newbies' First Interactions on Project Mailing Lists." In *Proc. of HICSS '11*, 1-10.
12. Katz, J. "Luring the Lurkers." 1998. <http://news.slashdot.org/story/98/12/28/1745252/Luring-the-Lurkers>
13. Krogh, G., Spaeth, S. and Lakhani, K. "Community, joining, and specialization in open source software innovation: a case study." *Research Policy*, 32, 7 (2003), 1217-1241.
14. Lakhani, K.R. and Wolf, R.G. "The Boston Consulting Group Hacker Survey." 2002. <ftp3.au.freebsd.org/pub/linux.conf.au/2003/papers/Hemos/Hemos.pdf>.
15. Lampe, C. and Johnston, E. Follow the (Slash) dot: Effects of Feedback on New Members in an Online Community. *Proc. of the 2005 Int. Conf. on Supporting Group Work*, 2005, 11-20.
16. Mason, B. Issues in virtual ethnography. *Proc. of Ethnographic Studies in Real and Virtual Environments: Inhabited Information Spaces and Connected Communities*. 1999, 61-69.
17. "MBOX Documentation." Qmail Mirror Selection. 1998. <http://www.qmail.org/man/man5/mbox.html>.
18. Nafus, D., Leach, J. and Krieger, B. "Deliverable D16: Gender: Integrated Report of Findings." Free/Libre/Open Source Software: Policy Support. <http://www.flosspols.org/>.
19. U.S. Department of Education, National Center for Education Statistics. *The Condition of Education* 2011. NCES, 2011, Table A-26-2. <http://nces.ed.gov/fastfacts/display.asp?id=72>.
20. "About:Fact Sheet." National Center for Women and Information Technology. <http://www.ncwit.org/about.factsheet.html>.

21. Nonnecke, B. and Preece, J. "Lurker Demographics: Counting the Silent." In CHI 2000, ACM Press (2000) 73-80.
22. Raymond, E.S. *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Beijing: O'Reilly, 1999.
23. Robert, K. "Standing out in a Crowd." Keynote Presentation, OSCON 2002, 2002. (Dreamwidth) <http://www.oscon.com/oscon2009/public/schedule/detail/10173>
24. Robles, G., Scheider, H., Tretkowski, I. and Webers, N. "Who Is Doing It? A research on Libre Software developers." 2001. <http://widi.berlios.de/paper/study.html>
25. Soroka, V., Jacovi, M. and Ur, S. We can see you: a study of the community's invisible people through ReachOut. In Huysman, M., Wenger, E. and V. Wulf (eds.), *Proc. of Int. Conf. on Communities and Technologies*, 2003. Kluwer Academic Publishers, 2003, 65-79.
26. "Users & Documentation." Parrot VM. Web. 21 Aug. 2011. <<http://parrot.org/dev/docs/user>>.
27. Wheeler, D. "Why Open Source Software/Free Software (OSS/FS, FOSS, or FLOSS)? Look at the Numbers!" April 16, 2007. http://www.dwheeler.com/oss_fs_why.html
28. "Women, Minorities, and Persons with Disabilities in Science and Engineering." U.S. National Science Foundation, 2011. <http://www.nsf.gov/statistics/wmpd/>
29. Ye, Y., and Kishida, K. Toward an Understanding of the Motivation Open Source Software Developers. In *Proc. ICSE 2003*, 419-429.
30. "Yum Package Manager." <http://yum.baseurl.org/>.