Where Are Females in OSS Projects? Socio Technical Interactions

Ikram El Asri[(✉)] and Noureddine Kerzazi

National Higher School for Computer Science and System Analysis (ENSIAS),
Mohammed V University, Rabat, Morocco
{ikram.asri,n.kerzazi}@um5s.net.ma

Abstract. Recent researches provide evidence that women are underrepresented in the field of computer science. It has been reported that less than 10% of Open Source Software (OSS) contributors in GitHub are women. Although related qualitative and quantitative studies point out the gender gap, the technical and social interaction of females within OSS still remain unexplored and largely misunderstood. As a first step towards proposing articulated actions towards diversity and inclusion, we need first to explore the gender gap in terms of activities and interactions. Thus, we propose to answer the questions: *where are females in OSS projects? How they evolve? and How they contribute to the sustainability of the OSS social capital?*. We particularly focus on building socio-technical networks and analyze them to explain how females contribute and interact in practice. We reflect on interactions' graphs and examine through a preliminary study, using data from six OSS projects, possible links between existing findings and the directions we suggest for more gender diversity. We found that females are extremely underrepresented within OSS communities, but when they participate they are productive just as males, they evolve following relatively the same patterns than males and remain more involved in projects than males.

Keywords: Gender · Diversity · Open Source Software ·
Social Network Analysis · Socio-Technical interactions

1 Introduction

Women are underrepresented in the software development industry and particularly within OSS communities [1–5]. In the software industry, females only account for 21% of the whole software development workforce and earn on average $22,251 less than their male peers [6]. In GitHub[1], a well-known open-source platform, the annual diversity report on OSS communities reported that the overall number of females decreased from 37% in 2017 to 33% in 2018. Repository data from GitHub projects provide an interesting source of information which, under appropriate analysis, may reveal facts related to females' activities and their dynamic interactions in a collaborative development environment. To this end, we undertake a socio-technical analysis

---

[1] https://github.com/about/diversity.

to get more insights into females' involvement in OSS projects. More specifically, we seek to answer the questions: Where Are Females in OSS Projects? How They Evolve? And How they contribute to the sustainability of the OSS social capital?

Why Diversity Matters? Diversity in the workplace including diversity of gender, ethnicity, and even religion has been shown to improve retention and reduce the costs associated with employee turnover [4, 7–9]. In a diverse workplace, employees are more likely to remain loyal when they feel respected and valued for their unique contribution [10]. That said, while GitHub is supposed to be a meritocracy based platform and free of gender barriers [11], the situation of females involvement is below the expectations [6] which limits the workforce required for communities growth. To benefit from gender diversity, OSS communities should lower barriers for female developers' self-guided personal development, which is crucial for them to achieve their own technical distinctions.

In order to support gender diversity within GitHub communities, we should understand how members, especially women, interact in these community environments. Understanding social interaction of female's collaboration could reveal insights not only about the structure of collaboration but also how productive are cross-gender collaboration: we know who contributed on what and when by examining historical data from GitHub. From there, we can build a social network of contributors (for both males and females) who have collaborated on the same source code files. Social Network Analysis (SNA) provides a class of metrics known as "centrality" metrics that provide a useful abstraction of the gender interaction information. Thus, we analyze and compare activity data (i.e., #commits, #comments, #fileModified, and code churn) between genders. Then we analyze SNA metrics of interactions (i.e., Female-Female, F-M, and M-M).

In particular, we answer the three following research questions:

RQ1: Where are females in OSS projects?
RQ2: How the top females' performers evolved over time?
RQ3: What is the role of females in building a sustainable collaborative social capital?

Paper Structure. We first discuss related work in Sect. 2. Then we present our methodology in Sect. 3 including data collection and gender detection approach. Section 4 introduces the basic terminology and concepts needed to comprehend how we built social networks, while Sect. 5 offers an overview of our findings along with discussion. In Sect. 6 we discuss possible threats to validity. Finally, Sect. 7 concludes and outlooks future work.

2   Related Work

The study by Vasilescu et al. [4] is the one that shares objectives closest to ours. The authors acknowledge the importance of gender diversity and explored how diverse are online teams with respect to gender and tenure based on the findings of a survey of GitHub contributors. Authors point out that when forming or recruiting a software

team, increased gender and tenure diversity are associated with greater productivity. Similarly, James et al. [12] studied the perception, performance, team dynamics, and opportunities using a survey targeting software professionals and reported that there is no significant difference between genders. Medenz et al. [13] investigated a gender inclusive method and how it can help increase gender inclusiveness in the tools that are used by OSS communities.

Wang et al. [11] discussed the existence of the competence-confidence gap. Authors developed a theoretical explanation for female developers' low rate of initiating a pull request in OSS projects arguing that it is not easy for female developers to directly translate competence to confidence [11]. Similarly, Terrell [14] compared acceptance rates of contributions from men and women in an open-source software community and finds that, in overall, women's contributions tend to be accepted more often than men's - but when a woman's gender is identifiable, her Pull requests are rejected more often.

In this study, we investigate the socio-technical interactions of females in OSS communities using SNA. We leverage on the historical data of six projects from GitHub. The benefit of assessing females' position in the overall collaborative networks is that it does not require qualitative information that can only be provided by contributors, such as questions about what they have done before. Analyzing networks of collaboration such as file co-edition, comments, pull requests offer opportunities for understanding the value provided by women in software development and help other researchers recommending improvements.

3   Methodology

3.1   Data Collection

The primary goal of our study is to understand how females interact and evolve within OSS communities. To this end, we performed a preliminary analysis of socio-technical interactions on publicly available historical data from six open source projects, mined from GitHub: *Angular.js, Moby, Rails, Tensorflow, Django, Elasticsearch*. We selected these well-known OSS projects because they are long-lived with more than a thousand contributors, and under diverse programming languages. Table 1 shows descriptive statistics regarding the studied projects.

In order to understand the female's distribution along with their activities within OSS projects, we first extracted information for each commit including the login of contributors, Timestamp of the commit, number of files modified, and code churn information (i.e., quantification of the commit size). Next, we used a Rest GitHub API[2] to extract detailed information of the contributor for each commit. For instance, the HTTP GET request "https://api.github.com/users/Narretz" sends back available details of the account including contributor's name "*Martin Staffa*". However, users registered in GitHub can choose to disclose their names in their profiles, so that out of the six projects studied, 1059 contributor names were missing. To discover as many of the

---

missing names as possible, the GitHub API provides access to user's public events which list all public events performed by a user on the site. This list provides another way to view the personal information associated with GitHub data. For instance, calling "https://api.github.com/users/bumbu/events/public" returns a Json object that includes a full name associated with the user and commits 'Alex Bumbu'. Using this method, we were able to detect 155 names from the missing list since GitHub users can upgrade their privacy settings to hide personal information[3]. We verified that less than 10% of contributors' names are missing for each project. Unknown names in the profile cannot be used to identify genders, we ignored them in our analysis.

Table 1. Quantification of activities by gender. (1) unknown gender is filtered out; (2) numbers are normalized by the number of contributors for each gender.

Project	Lang	Contrib	Commits	%Females	%Males	Commits		Comments		Edited Files		Code Churn		p_value (F vs M commits)
						F(%)	M(%)	F(%)	M(%)	F(%)	M(%)	F(%)	M(%)	
Angular.js	JavaScript	1601	8897	3.4	76.1	2.79	5.86	0.18	0.94	7.62	20.37	291.5	1162.13	0.17
Moby	Go	1824	36007	3.5	88.1	20.08*	18.73*	0.01	0.19	131.6	93.69	7051.97	5612.75	0.2
Rails	Ruby	3723	70762	4.2	91	8.07	19.21	0.73	4.63	29.62	70.25	77.94	1912.8	0.79
Django	Python	1672	26283	5.3	88	5.51	16.43	0.11	0.84	17.34	70.05	271.48	3467.39	0.09
Elasticsearch	Java	1127	42702	4.2	86.1	25.45	42.24	0.04	0.56	166.08	481.9	6839.1	27271.75	0.43
TensorFlow	C++	1735	44132	5.8	80.7	17.92	25.65	0.09	0.21	131.23	258.34	19601.38	43082.54	0.14

3.2 Gender Detection

GitHub does not store information about contributors' gender identification. There are few approaches to automatically infer a contributor's gender [4, 14, 15]. Vasilescu et al. [4] used first name and country to infer a user's gender. Terrell et al. [14] suggested a method using email to link a user to her Linkedin profiles which contains rich information. Although this method shows a high precision, the recall is not good, especially for female developers from East Asian countries where the usage of Linkedin is not that high. However, these heuristics have not been empirically validated with GitHub data. To alleviate issues related to previous approaches, we used the commercial tool called Namsor API[4]. Table 1 shows a comparison of females and males percentage distributions.

4 Social Network Representation

Social Network Analysis (SNA) is the process of investigating social structures through the use of networks and graph theory [16]. In network analysis, we have nodes (i.e., contributors) which represent vertices of a graph and connections which represent edges (i.e., a type of interaction). An interaction occurs when two developers modify the same source code file or comment on the same topic.

---

[3] https://medium.freecodecamp.org/github-privacy-101-how-to-remove-personal-emails-from-your-public-repos-58347b06a508.

[4] http:/blog.namsor.com/api.

SNA Metrics. Metrics that measure a node's direct connections to other nodes are connectivity metrics. The first metric is the **density** of a network which captures the number of actual connections between members divided by the number of possible connections, as depicted in Fig. 1. Density values range from [0 to 1], a higher density indicates that network nodes have tighter connections with each other. Second, connectivity metric is the **degree centrality** of a node which represents the number of connections incident on a node. Each time two contributors change the same file or comment on the same topic we have one non-weighted connection (i.e., a link). In contrast, a node is considered disconnected if it has no edges with other nodes meaning no interactions with other contributors. Centrality metrics quantify how closely contributors are indirectly connected to other contributors in the network. We consider two metrics: closeness and betweenness. **Closeness** stands for the average number of steps required to go from the current node to all other nodes. Closeness measures the distance between each participant and all other participants (i.e., is a participant connected directly to all other participants, or would information need to pass through several other participants to reach that individual?). **Betweenness** stands for the average number of shortest paths between pairs of other nodes that run through the node. Betweenness is used to measure the extent to which a node lies between other nodes. More shortest paths run through a node, more likely important this node is. **Clustering Coefficient** is calculated as the probability that any two neighbors of the current node are connected. In our study, the clustering coefficient measures the collective collaboration of contributors from different genders.

We used SNA metrics to explore diversity of interactions of Female-Female (**F-F**), Female-Male (**F-M**), and Male-Male (**M-M**) using files co-edition networks from six GitHub projects.

5   Results

RQ1. Where Are Females in OSS Projects?

Motivation. Social capital of open source software needs gender diversity among other diversities (i.e., cultural, ethnic, etc.), to be sustainable [17]. Gender diversity demands coordinated and delicate interactions [4]. Unfortunately, little is known about how females behave and interact within OSS communities, comprehensive view is rarely discussed at least to understand what would be specific for females to ease their integration within OSS communities.

Approach. We first extract historical data of Commits, Comments, and Pull Requests from GitHub with their respective authors similarly to the approach provided by Joblin et al. [18]. After identifying the gender of each contributor, we focused on three groups of interactions cross-gender F-F, M-M, and F-M. Thus, for each project, we built a social network, along with subnetworks regarding the three groups, related to contributors' interactions for the six studied projects. More precisely, by interactions we mean the relation by which two or more contributors collaborate with each other through an activity such as working on the same source code file (i.e., co-edition) or

commenting on the same commit. We tracked back this information by looking at the version change history and other elements provided by the GitHub platform.

Results. Figure 1 shows social networks related to contributors' interactions for the six studied projects. Nodes represented in red color point out males and yellow ones highlight the position of females within the overall network. Visually, as shown in Fig. 1, females are distributed throughout the overall network. They are positioned within the Core as well as Peripheral members and are interacting with both genders. Furthermore, to examine interactions regarding gender, we calculate the SNA metrics of three sub-networks Fig. 2: *(i)* interactions between females F-F, *(ii)* interactions between males M-M, and *(iii)* interactions with the opposite gender F-M. Table 2 reports the averages distribution of the most important SNA metrics cross-networks.
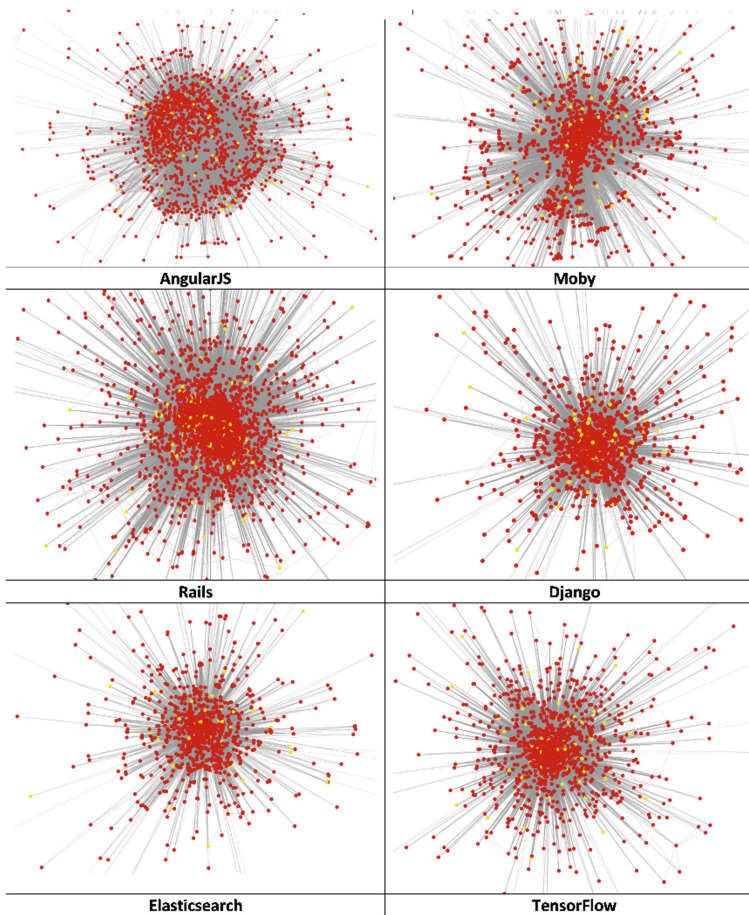

Fig. 1. Social position of females within the socio-technical network (females are represented in yellow). (Color figure online)

F-F interaction sub network
#node: 54
Density :0.05

M-M interaction sub network
#node: 1394
Density:0.05

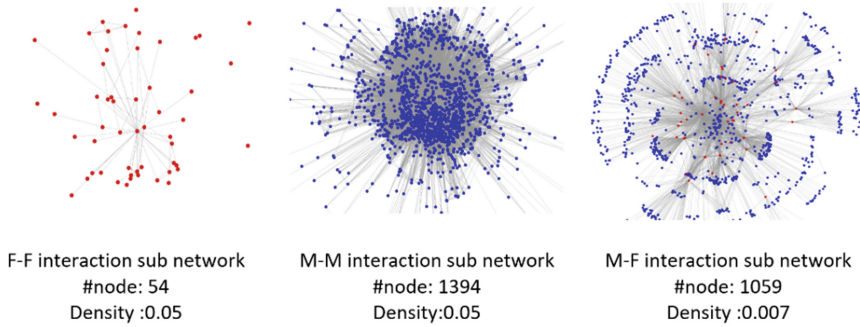M-F interaction sub network
#node: 1059
Density :0.007

Fig. 2. Sub networks representing cross-gender interactions.

For all projects, F-F interactions are as dense as M-M and F-M interactions are less dense meaning that there are more interactions between contributors within the same gender group than the opposite gender. For example, within Angular.JS project, the two networks F-F and M-M have the same density equal to 0.05, however, the density of cross-gender interactions network M-F is very low (=0.007).

The *Avg. degree* metric for the sub-networks F-F is extremely low compared to M-M meaning that females are less directly connected to other females (*2.7* for Angular.js) and also less connected to males (7.05). We hypothesize that this result is skewed because of the unbalanced numbers of contributors for each gender.

The *Avg.ClusCoef* related to the sub-network M-M is much higher than F-F meaning that there are established groups of collaboration (for Angular.js, Avg.ClusCoef = *0.89* between males (M-M) against *0.46* for F-F) co-editing the same components for a long time. While Avg.ClusCoef value close to 0 for the sub-networks M-F suggesting that there are few cross-gender interactions.

Closeness metric is roughly similar cross-gender meaning that gender has no effect on the proximity of members with other members of the community and how fast the members interact with each other (i.e., fewer steps to reach other nodes).

Betweenness metric has a very small value suggesting that nodes are not that much connected (value = 1 means a hub node and value = 0 means a disconnected node). For instance, in Angular.js project, the average Betweenness for the network F-F equal to 0.02, while it is equal to 6e-4 for the sub-network M-M.

Table 2. Summary of the most important SNA metrics using sub-networks related to gender.

Project	Density			Avg.degree			Avg.ClustCoef			Avg.ClosCentrality			Avg.BetwCentrality		
	F-F	M-M	M-F	F-F	M-M	M-F	F-F	M-M	M-F	F-F	M-M	M-F	F-F	M-M	M-F
Angular.js	0.05	0.05	0.007	2.7	69.58	7.05	0.46	0.89	0	0.41	0.51	0.37	0.02	6 E-04	1 E-03
Moby	0.07	0.05	0.008	6.24	85.03	11.34	0.54	0.87	0	0.41	0.51	0.41	0.01	6 E-04	1 E-03
Rails	0.05	0.06	0.006	8.34	230.67	19.83	0.63	0.85	0	0.42	0.51	0.42	E-03	2 E-04	4 E-04
Django	0.09	0.1	0.014	8.5	161.45	19.37	0.65	0.84	0	0.42	0.53	0.4	0.013	5 E-04	1 E-03
Elasticsearch	0.04	0.05	0.008	2.08	49.31	5.85	0.29	0.89	0	0.31	0.5	0.4	0.029	1 E-03	2 E-03
TensorFlow	0.08	0.07	0.014	7.91	117.38	17.95	0.53	0.89	0	0.42	0.52	0.39	0.01	6 E-04	1 E-03

Also, as one can see in the subnetwork M-F in Fig. 2, females can play a broker role ensuring the communication between sub groups, which can decrease the centralization of the OSS community and increase communication between Core (i.e., experts) and Peripheral members (i.e., casual contributors) [19].

> **Where are females in OSS projects?**
> Although females are underrepresented within OSS communities, we found them in the Core teams as well as in Peripheral. Women are more likely than men to interact with other contributors. However, the density of cross-gender interactions network M-F is very low (=0.007) with less established groups of females (0.46) compared to males (0.89).

RQ2. How the top females' performers evolved over time?

Motivation. A common problem that OSS communities face is the high instability of their contributors [20]. This problem is even amplified by the fact that females are underrepresented, their permanence is also an open issue. We sought to gain more understanding of the evolement of female contributors over time and to learn from the experience of the top performers.

Approach. For each project, we have retrieved the third quartile (i.e., upper quartile that has the top 25% above it) of the female developers that contribute the most according to the: number of commits, comments, and Pull Requests. From it, we calculate the performance for each contributor according to the following formula.

$$Performance(C_i) = \frac{1}{Total\_activity}\left[date\_lastActivity(C_i) - date\_firstActivity(C_i)\right]$$

Results. Figure 3 shows the distribution of performance between males and females. We performed Mann-Whitney-Wilcoxon statistical tests, with a confidence level of alpha = 0.05. We found that for all projects except Moby the tests were not statistically significant ($p > 0.05$) suggesting that there is no statistical difference of the performance cross-gender except for the project Moby as reported in Table 3.
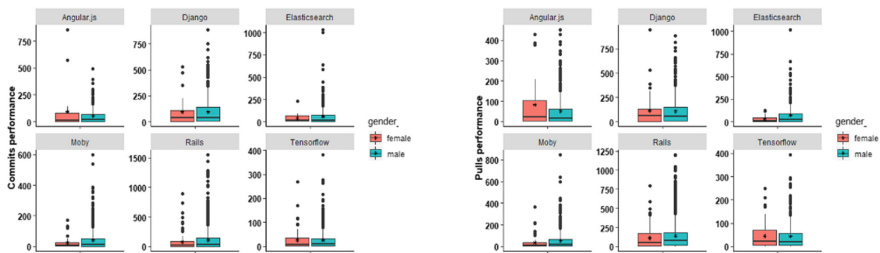

Fig. 3. Comparison of the gender performance

Table 3. ManWhitney significance test

Project	p-value(commits)	p-value(PullRequests)
Angular.js	0.74	0.23
Moby	0.37	0.14
Rails	**0.02**	0.2
Django	0.49	0.6
Elasticsearch	0.78	0.11
Tensorflow	0.72	0.73

> **How the top females' performers evolved over time?**
> Females evolve just as males within OSS projects. There is no statistical difference of the performance cross-gender except for the project Moby.

RQ3. What is the role of females in building a sustainable collaborative social capital?

Motivation. Sustained participation in crucial for successful OSS project [19]. We expect that the more often individuals participate in OSS, the higher their chance of prolonged engagement.

Results. Table 1 summarizes the involvement of females in OSS projects. The percentage of females range from *3.4%* to *5.8%* which is drastically less than the average numbers presented in the annual report of GitHub [6]. When normalizing these numbers by genders, we noticed that women contribute slightly less than men except for the Moby project where women outperform (*20.08%* vs. *18.73%* for men). Apparently, in the Angular.js project, the productivity of men is two times higher than women in terms of commits (*2.79%* vs. *5.86%*). Similarly, we report on the difference of productivity for other projects (Moby = *0.9*; Rails = *2.3*; Django = *2.9*; ElasticSearch = *1.6*; TenserFlow = *1.4*).

Moreover, Table 4 shows the evolution of the activity related to both genders (due to lack of space, we report data only for Angular.js). Although OSS projects keep attracting both genders, the number of active contributors (i.e., committing and commenting) decreases proportionally for both genders.

Table 4. Activity evolution. Example of Angulat.Js project.

Year	Contributors		Commits		Edited Files		Code Churn		Comments	
	F	M	F	M	F	M	F	M	F	M
2010	0	6	0	6 → 581	6 → 2687	6 → 175340	0	2 → 4	0	0
2011	0	20	0	17 → 811	17 → 4095	17 → 249305	0	6 → 20	0	0
2012	2	97	↑ 2 → 2	84 → 754	2 → 4	84 → 2848	2 → 24	84 → 124359	0	18 → 32
2013	17	511	↑ 16 → 29	436 → 1503	16 → 16	436 → 4460	16 → 1309	436 → 209275	1 → 1	135 → 276
2014	34	965	↑ 20 → 72	517 → 2131	20 → 230	57 → 6111	20 → 4354	516 → 263323	3 → 4	166 → 525
2015	45	1179	↓14 → 34	270 → 960	14 → 50	270 → 3079	14 → 9914	270 → 138804	2 → 3	79 → 229
2016	47	1335	↓5 → 5	184 → 879	5 → 5	184 → 3343	5 → 14	184 → 351669	1 → 2	37 → 163
2017	52	1338	↑ 6 → 6	65 → 305	6 → 79	65 → 1254	5 → 68	65 → 75817	0	14 → 32
2018	54	1395	↓3 → 3	24 → 252	3 → 5	24 → 548	3 → 58	24 → 33280	0	17 → 35

> **What is the role of females in building a sustainable social capital?**
> Even if females are underrepresented in OSS communities, they contribute in building a sustainable social capital for open software.

6 Threats to Validity

This section discusses the threats that might have affected our findings, and how we have alleviated these threats.

Threats to *Internal Validity* concern alternative factors that could have influenced our results. This exploratory study might suffer from at least the similar threats that other studies using GitHub Data do [21] related to possible issues with data gathering and the missing of validation. We believe that using a large amount of data mitigates this threat. Another threat relates to the accuracy of gender prediction. Various approaches and tools to infer gender based on names and countries have been proposed in the literature [4, 22, 23]. Most of these tools rely on English names stored in databases. However, these heuristics have not been yet empirically validated with GitHub data. We tried three tools [4, 14, 15], and found that each has strengths and drawbacks. We used *Namsor*, a commercial tool with the assumption that it provide the most accurate results. We also assume the gender is a binary attribute.

Threats to *construct validity* consider the agreement between a theoretical concept and a specific measuring procedure. One threat considers the intrinsic blind spots when investigating consecutive snapshots of a dynamic social network over time-stamps (years). Indeed, there is no way to ensure that years are the "optimal" time-stamps to represent the snapshots of our social networks in order to capture the evolement and interaction between nodes (contributors). Therefore, tracking a community evolution in such a constraint can yield evolving observations in a non-consecutive way. Furthermore, some important events (e.g., major releases, vision changes, etc.) that communities may undergo across time-stamps will not be detected and will be difficult to predict.

Threats to *external validity* relate to the generalization of our findings. We only considered six OSS projects from GitHub. Thus, we cannot assume a generalization to all projects hosted on GitHub or other platforms such as GitLab and Bitbucket, even though there is no inherent reason why they would be biased.

7   Conclusion

We have used SNA approach to study gender diversity within six GitHub communities. We first examined the position of both genders within social networks and their interactions and found that females are spread over the overall network including Core and Peripheral members. Next, we tracked back the evolvement of the upper quartile of the female developers that contribute the most according to the: number of commits, comments, and Pull Request. We found that there is no statistical difference regarding the performance cross-gender except for the project Moby. Finally, even if females are underrepresented in OSS communities, they contribute in building a sustainable collaborative social capital for open software.

The results of this study motivate further fine-grained analysis of female's implication in OSS that will *(i)* investigate the quality of work in terms of introduced bugs or complexity; *(ii)* explore further socio-technical interactions of women to understand the value of their contributions; *(iii)* characterize projects attractiveness for women.

References

1. Feller, J., Fitzgerald, B.: A framework analysis of the open source software development paradigm (2000)
2. Nafus, D.: 'Patches don't have gender': what is not open in open source software. New Media Soc. **14**(4), 669–683 (2012)
3. Steinmacher, I., Chaves, A.P., Conte, T.U., et al.: Preliminary empirical identification of barriers faced by newcomers to Open Source Software projects, pp. 51–60 (2014)
4. Vasilescu, B., Posnett, D., Ray, B., et al.: Gender and tenure diversity in GitHub teams. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, pp. 3789–3798 (2015)
5. Filippova, A., Trainer, E., Herbsleb, J.D.: From diversity by numbers to diversity as process: supporting inclusiveness in software development teams with brainstorming, pp. 152–163 (2017). %@ 1538638681
6. Datausa.io. Demographics of Software Developers
7. Hui, J.S., Farnham, S.D.: Designing for inclusion: supporting gender diversity in independent innovation teams, pp. 71–85 (2016)
8. Imtiaz, N., Middleton, J., Chakraborty, J., et al.: Investigating the effects of gender bias on GitHub (2019)
9. Foucault, M., Palyart, M., Blanc, X., et al.: Impact of developer turnover on quality in open-source software, pp. 829–841 (2015)
10. O'Reilly Iii, C.A., Caldwell, D.F., Barnett, W.P.: Work group demography, social integration, and turnover. Adm. Sci. Q., 21–37 (1989)

11. Wang, Z., Wang, Y., Redmiles, D.: Competence-confidence gap. In: Proceedings of the 40th International Conference on Software Engineering Software Engineering in Society, ICSE-SEIS 2018, pp. 81–90 (2018)
12. James, T., Galster, M., Blincoe, K., et al.: What is the perception of female and male software professionals on performance, team dynamics and job satisfaction?: insights from the trenches, pp. 13–22 (2017)
13. Mendez, C., Sarma, A., Burnett, M.: Gender in open source software: what the tools tell, pp. 21–24 (2018)
14. Terrell, J., Kofink, A., Middleton, J., et al.: Gender differences and bias in open source: pull request acceptance of women versus men. PeerJ Comput. Sci. 3, e111 (2017)
15. Vasilescu, B., Capiluppi, A., Serebrenik, A.: Gender, representation and online participation: a quantitative study of stackoverflow, pp. 332–338 (2012)
16. Zhang, W., Nie, L., Jiang, H., et al.: Developer social networks in software engineering: construction, analysis, and applications. Sci. China Inf. Sci. 57(12), 1–23 (2014)
17. Dabbish, L., Stuart, C., Tsay, J., et al.: Social coding in GitHub: transparency and collaboration in an open software repository. In: The conference on Computer Supported Cooperative Work, Seattle, WA, USA, pp. 1277–1286 (2012)
18. Joblin, M., Apel, S., Mauerer, W.: Evolutionary trends of developer coordination: a network approach. Empirical Softw. Eng. 22(4), 2050–2094 (2016)
19. Toral, S.L., Martínez-Torres, M.R., Barrero, F.: Analysis of virtual communities supporting OSS projects using social network analysis. Inf. Softw. Technol. 52(3), 296–303 (2010)
20. Robles, G., Gonzalez-Barahona, J.M., Herraiz, I.: Evolution of the core team of developers in libre software projects. In: 2009 6th IEEE International Working Conference on Mining Software Repositories, pp. 167–170 (2009)
21. Kalliamvakou, E., Gousios, G., Blincoe, K., et al.: The promises and perils of mining GitHub, pp. 92–101 (2014)
22. Lin, B., Serebrenik, A.: Recognizing gender of stack overflow users, pp. 425–429 (2016)
23. Karimi, F., Wagner, C., Lemmerich, F., et al.: Inferring gender from names on the web: a comparative evaluation of gender detection methods, pp. 53–54 (2016)