


Gender Diversity on GitHub Issue Tracking: What's the Difference?

Estela Miranda Batista  [Universidade Federal de Viçosa | estela.batista@ufv.br]

Gláucia Braga e Silva  [Universidade Federal de Viçosa | glaucia@ufv.br]

Thais Regina de Moura Braga e Silva  [Universidade Federal de Viçosa | thais.braga@ufv.br]

Abstract This work analyzes female participation in communication on GitHub's Issue Tracking, based on thematic relevance of posted comments according to the developer's gender relative to other metrics, such as reputation, participation time on the platform, and number of reported issues. Data from 5 open source communities and 5 communities dedicated to women was analyzed. The results indicate that, on average, the relevance of comments made by women is similar to that of men. However, the study confirms other findings in literature that highlight low levels of female representativeness and participation in projects, with just 22% of comments posted by women and 16% of issues reported by them.

Keywords: *Gender on GitHub, Communication in Issue Tracking, Blau Index*

1 Introduction

In software development environments, teams can be formed by developers with different seniority levels and different diversity profiles. Diversity can be classified into three categories: demographic (gender, ethnicity, age), psychological (values, beliefs, and knowledge), and organizational (time of experience, occupation, hierarchical level) (Outão and Santos, 2022). Teams with greater social diversity can benefit from broader information, experience, and improved problem-solving skills, becoming more effective (Vasilescu et al., 2015). For some categories, there are still many challenges related to low team representativeness, such as gender diversity. Studies show low participation rates and lower permanence of women when compared to men (Vedres and Vasarhelyi, 2019). Zacchiroli (2021) analyzed 1.6 million commits and showed that men did about 92% of the code produced on code versioning platforms.

The low female participation in software projects is known in the literature and has been mainly evaluated in terms of accounting for code contributions and accepted pull requests (Rodriguez et al., 2021; Canedo et al., 2019). However, other aspects can be analyzed in order to investigate female participation in the composition of more diverse, more inclusive, and potentially more productive teams (Catolino et al., 2019). Among the possible aspects that can be analyzed, the participation of women in communication on software projects stands out. Despite the low representativeness and less permanence in projects, studies show that women act as problem mediators (Catolino et al., 2019), being key players in the communication process.

This work, which extends a previous research (Batista et al., 2022), evaluates the impact of gender diversity on the quality of communication in issue tracking environments, using the Thematic Relevance metric of posted comments (Neto and Silva, 2018). This metric is capable of representing how relevant a text is in regards to the theme of a given discussion (Machado et al., 2019). We conduct this analysis along with other metrics such as the number of issues created and the number of posted comments, based on other attributes that

characterize the developer profile, such as gender, platform participation time and reputation. The study was guided with the following research questions:

- **RQ1.** What is the difference in the participation of men and women in terms of issues and associated comments?
- **RQ2.** Is there a difference between the relevance of comments posted by men and women?
- **RQ3.** Is there a relationship between comment relevance and platform participation time?
- **RQ4.** Is there a relationship between comment relevance and author's reputation?

To answer the listed research questions, we evaluated communication data extracted from open source project repositories on the GitHub platform, in search of comparative analyzes between the participation of men and women in the context of issue tracking. GitHub was used because it is the largest open source community, with more than 61 million software repositories created by more than 16 million people registered on the platform¹, while also being increasingly used among researchers as a source for data mining (Saadat et al., 2020). To compare the participation of women and men in these environments, we extracted data from two types of open source communities: open communities and communities dedicated to women. Analyzing data from different communities is still an uncommon issue to be explored in literature (Outão and Santos, 2022), which is important for understanding the interaction of members and how different communities can positively or negatively affect a developer.

As a result, we present relevant indications about the teams' communication process and the possible differences in terms of contributions depending on the developer's gender. We emphasize that the term gender is treated here based on how users are identified on the platform, that is, the possible names assigned, which can assume masculine or feminine values so that no differentiation is made between cisgender or transgender people. Since this information is not recorded directly on the GitHub platform, it is not a study object of this research.

¹<https://octoverse.github.com/>

The remainder of the paper is organized as follows: Section 2 presents information on Issue Tracking communication and the Thematic Relevance concepts applied to comments. In Section 3, related works are presented. Section 4 shows materials and methods used in this work, while results are discussed in Section 5. Threats to validity of the work are described in Section 6, and in Section 7 the final remarks are presented.

2 Communication in Issue Tracking

Throughout software development, project members and users raise several doubts and questions about improvements. These topics are commonly concentrated in discussion threads, represented by *issues* in the GitHub issue tracking system. According to Bertram et al. (2010), issue tracking repositories are knowledge repositories, which concentrate much of the communication and collaboration of projects, producing useful and relevant knowledge.

Automated analysis of this data can reveal useful process quality indicators regarding the participation of members in project discussions, such as key developers, comment relevance, conflicts, representativeness, and participation of newcomers and minorities. To assist this analysis process, certain metrics were used, such as user reputation, years of user participation on the platform, and comment quality.

2.1 Issue Tracking on GitHub

In issue tracking environments, communication occurs as issues – representing bugs, improvement suggestions, and new requirements – are reported by project members or external collaborators. After being reported, issues can also be resolved or discussed by any team member or external collaborators. During the discussion in search for solutions, new commits can be created, and consequently pull requests can be evaluated by project managers, or by someone else responsible for the repository.

Issues are made up of some mandatory fields, and others that, depending on the project, do not need to be filled in. The mandatory fields are the title and a description. Optional fields include labels, such as "bug" or "question", and who the issue is assigned to. Finally, there are fields automatically filled by the platform, such as the author, creation date, and status (open or closed). The fields related to an issue can be seen in Figure 1².

The title and description fields are the main source of textual data, since they represent the topic of the discussion held in the issue through subsequently posted comments. Thus, these fields will be explored in the context of this work.

2.2 Thematic Relevance in Issue Tracking

The thematic relevance metric was proposed by Azevedo (2011) to analyze the relevance of comments posted in educational discussion forums and later adapted for the issue tracking environment by Neto and Silva (2018). We aim to

use thematic relevance in the context of issue tracking to identify whether the comment text is related to its respective issue topic, in order to assess its impact on each issue resolution. With this metric, it is also possible to investigate whether there is a correlation with other attributes linked to the developer profile, such as gender, reputation, and project time.

The calculation of this metric involves counting the number of relevant concepts used in the comment text that correspond to concepts in the issue text, composed of title and description. In addition to accounting for relevant concepts, the equation also considers their frequency in the text and relationship with other terms. Thus, the more related concepts, the greater the relevance of the comment. Equation 1, adapted from (Neto and Silva, 2018), presents the calculation of thematic relevance.

$$TR = \max\{S_{CI}, S_{CD}, S_{CC}\} \quad (1)$$

Where TR is the Thematic Relevance of a comment; S_{CI} , the similarity between the comment and the issue, represented by its title and description; S_{CD} , the similarity between the comment and the discussion, which takes into account the issue title, its description and the previous comment to the one analyzed; and S_{CC} , the similarity between the comment and the previous comment, if any. The value of TR is given by the highest similarity value found.

To calculate the presented similarities, two techniques can be used: graphs and cosines. As presented by Machado et al. (2019), in the graph technique, comment text is represented in a graph, in which the most relevant terms constitute the vertices and the edges connect the terms that appear together in the text. This graph is then compared with another graph, constructed in a similar way, but based on the issue text. For comparison between the graphs, the similarity is calculated from the correspondence between the terms. The cosine technique proposes that each text is vectorized so that an angular analysis can be performed between the textual contents (Brandão et al., 2022). Thus, when comparing, e.g., a comment vector with an issue vector, the cosine of the angle between these two vectors representing an interaction is calculated, and this value is assigned as the similarity.

3 Related Works

Low female participation in software development projects has been reported throughout literature. Izquierdo et al. (2019) analyzed more than 7000 user profiles on the GitHub platform, in which only 7% of project committee members and 8% of leaders were identified as women. Seeking to present new indicators on the low female participation in software development, we seek to analyze the quantity and quality of the content of communication in issue tracking.

With regard to the analysis of issue tracking communication data, works in the literature have used text mining techniques on the textual content of issues. Ortu et al. (2018) analyzed communication data in software projects in order to assess the impact of certain factors on the time taken to resolve an issue. Neto and Silva (2018) also analyzed issue tracking data in order to identify key developers with metrics such as the thematic relevance of comments, the number

²Adapted from <https://docs.github.com/pt/issues/tracking-your-work-with-issues/planning-and-tracking-work-for-your-team-or-project>

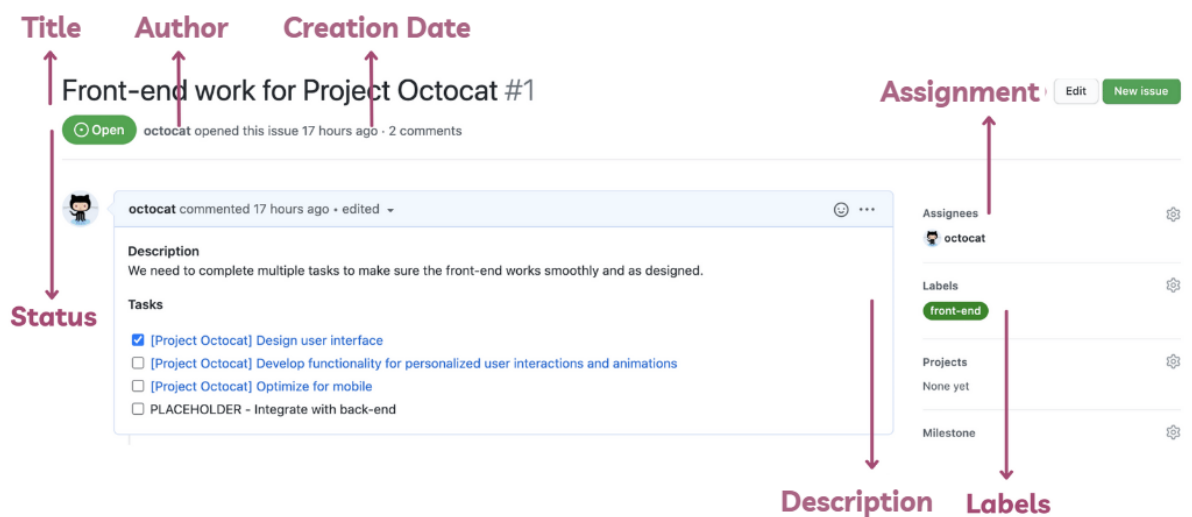


Figure 1. Example issue on GitHub.

of issues reported, and the number of comments posted on issues. However, despite bringing relevant investigations to software development communication, the works presented do not take gender issues into account.

Noei and Lyons (2022), in turn, evaluated around 700,000 app review comments on the Google Play Store platform, in which it was possible to acknowledge that women made fewer comments than men. In addition, the authors also observed most of the comments posted by women are positive, as they show praise for the applications, while those of men tend to bring criticism and suggestions for improvements. This means that adjustments and new versions developed as a result of comments mostly reflect the male opinion.

Considering female participation in the open source universe, Singh (2019) analyzed 355 project websites, and the results show that less than 5% of these communities had spaces dedicated to women. In this work, we use environments dedicated to women for the sake of comparison with low diversity environments, indicated by the Blau index (Blau, 1977), in order to verify whether women are more active in terms of communication in dedicated environments. The Blau index uses the percentage of individuals in a given category of a total population to determine the diversity of the population. This index will be detailed in the section 4.

Investigating other factors that can be analyzed in relation to the participation of women and men during software development, Steinmacher et al. (2012) evaluated the participation of beginner developers in an open source project. According to the authors, due to the lack of information and guidance during the first steps in a software project, newcomer developers generally post more questions and request help in their tasks. Qiu et al. (2019) conducted interviews with potential developers on open source projects and found that female newcomers are more cautious about joining open source projects. Furthermore, it was observed that men were considered more active contributors and more excited about joining co-ed projects.

Another important analysis is that of data regarding software development during the COVID-19 pandemic. Araujo et al. (2022) carried out a field survey with Brazilian developers in order to understand the disadvantages women have during remote work related to health, pressure, multitasking,

and exhaustion. Seeking to present new indicators on the participation of beginner developers and the disadvantages experienced by women during the COVID-19 pandemic, we analyze the quantity and quality of communication content in this context in issue tracking.

Finally, Table 1 summarizes the topics presented above, in relation to the themes proposed in this work.

4 Materials and Methods

To analyze the issue tracking communication data according to the developer gender, we performed three steps: defining the metrics, calculate thematic relevance of comments and extracting the data.

4.1 Definition and Automation of Metrics

To conduct the analysis, six gender-dependent metrics were used, called Communication Metrics (CM).

- **CM01** Number of Reported Issues - sum of issues reported by each developer;
- **CM02** Number of Posted Comments - sum of comments posted on an issue by each developer;
- **CM03** Platform Participation Time - difference between the entry and current years;
- **CM04** Developer Reputation - value assigned using developer connections, considered a social attribute;
- **CM05** Teams Gender Diversity Index - percentage of individuals in both categories (male or female);
- **CM06** Thematic Relevance of Comments - importance of each comment regarding an issue discussion.

To obtain metrics CM01 to CM04, APIs and tools from the GitHub platform were used. For metrics CM05 and CM06, an application was developed in Python, as described throughout this section.

Metrics CM01 and CM02 were calculated according to the gender of the developer. After identifying the gender of each developer through the NamSor³ tool, data was stored

³<https://namsor.app/>

Table 1. Comparison between related works.

	Gender Diversity	Issues	Comments	Communities	Thematic Relevance	Newcomers	Pandemic
Izquierdo et al. (2019)	X						
Ortu et al. (2018)		X					
Neto and Silva (2018)		X	X		X		
Noei and Lyons (2022)	X		X				
Singh (2019)	X			X			
Araujo et al. (2022)	X						X
Steinmacher et al. (2012)						X	
Qiu et al. (2019)	X					X	
This work	X	X	X	X	X	X	X

in a CSV file, and using filters in the Jupyter Notebook⁴, such calculations were performed. The CM03 metric was calculated according to the data provided by the GitHub API, subtracting the entry year of the developer from the current year of this work (2022).

For the CM04 metric, since the GitHub platform does not automatically determine developer reputation, the GitScore⁵ tool was used. The tool has three fields: GitScore, Reputation, and Contribution, with the second being used as the metric value. This field takes into account the connections of the developer, considered a social attribute, to complement other metrics regarding developer contribution, such as the number of comments.

In regards to the gender diversity of teams (CM05), the Blau index was used as a diversity metric (Blau, 1977). The diversity metric was used in order to compare the results of metrics applied to the communication data for each evaluated community segments. Equation 2 presents the formula for the Blau index, which calculates, from a total of N categories, the percentage P of individuals in each category i . In this work, we consider $N = 2$, since the categories are male and female. The index varies from 0 to 0.5, with 0.5 being the balance in the number of individuals in the categories.

$$Blau = 1 - \sum_{i=1}^N P_i^2 \quad (2)$$

After extracting data from a given community, the names and genders of the participating developers were stored in a CSV file. From this file, the number of female and male developers was counted using Jupyter Notebook filters, and used as input to a simple Python implementation of the Blau index, thus calculating the diversity level of that community.

Since on the GitHub issue tracking platform it is not mandatory to identify the gender of developers, and the previously presented metrics make comparisons in relation to gender, it was necessary to estimate such information. Zolduarrati and Licorish (2021) conducted a study to qualify the best gender prediction tools. Among the tools presented, after an evaluation for the context of this work, the NamSor tool was chosen, due to its simplicity of use and results quality. The tool takes as input the first and last name of a person and returns as output the probability of that person to be female or male, and so we assign their gender according to the highest

percentage. In addition, the tool supports several languages, such as Portuguese, English, Japanese and Chinese.

4.2 Thematic Relevance of Comments

To assess the quality of comments, an automated application for the thematic relevance metric was developed (Section 2.2). For simplification and performance improvement purposes, the metric implementation was adapted with regard to the similarity calculation between the terms of a comment and its reference text using the cosine technique. In the work of Neto and Silva (2018), the similarity calculation was performed through graphs, with an external dependency on a graph generation tool. However, this dependency caused performance issues due to the high number of requests. This replacement provided a significant performance improvement, without causing changes in terms of the quality of generated results, as detailed in Section 4.2.1. The cosine technique was also chosen due to its simplicity of application and also because it presents a high precision in relation to human classification, as observed by Medeiros et al. (2014).

While using the formula of Equation 1, presented in Section 2.2, we observed the comment similarity in relation to its previous comment (S_{CC}) assumed the value zero in several instances, showing to be irrelevant for calculating the thematic relevance. Thus, after some tests, it was possible to notice the comment similarity in relation to the discussion (S_{CD}) and the comment similarity in relation to the issue (S_{CI}) was sufficient to compose the thematic relevance calculation equation. Also, we opted for the arithmetic mean of S_{CI} and S_{CD} , as shown in Equation 3.

$$TR = \frac{S_{CI} + S_{CD}}{2} \quad (3)$$

4.2.1 Validation of Metric Adaptation

To validate the change and ensure reliability of the results, a validation procedure was performed, comparing the original results of Neto and Silva (2018) with the updated version by a manual evaluation conducted by specialists. The group was comprised of 12 volunteers in total, including 3 software engineers; 2 domain experts, i.e., people who participated in the issues; and 7 developers, with varying levels of expertise.

For the evaluation by specialists, a survey was elaborated to be answered for each issue, in which each specialist assigned a relevance from 0 to 4 for each comment in that issue.

⁴<https://jupyter.org/>

⁵<http://www.git-score.com>

Although the relevance is calculated in the range of 0 to 1, the integer interval of 0 to 4 was used to simplify the assignment by the specialists and avoid possible mistakes with decimal values. The values were later normalized to the same scale. In total, 59 comments were evaluated along 12 issues. Each issue was analyzed by two experts, so that each expert assigned relevant values to each of the comments registered for that issue. Along with the experts' evaluation, an average was calculated with the relevance assigned by the metric development team. Table 2 presents an example of the evaluation format of a given set of comments, in which the final relevance corresponds to the arithmetic mean between the relevance values defined by the specialists and the average relevance assigned by the development team.

Table 2. Example of manual assessment of comments thematic relevance on an issue

Comment	Average	Specialist1	Specialist2	Relevance
1	3.8	3	4	3.6
2	2.0	2	2	2.0
3	4.0	4	3	3.7
4	1.0	1	0	0.7

To verify the accuracy of the adapted metric approach, in comparison with the experts classification and with the original Neto and Silva (2018) version, the MAE (Mean Absolute Error) and MSE (Mean Squared Error) metrics were used. The MAE (Equation 4) represents the average of the absolute difference between actual and predicted values in the data set. The equation considers the final relevance value of the experts (y) and the automatically calculated values (\hat{y}), either by the original approach or by the adapted approach. Therefore, the calculation was performed twice, once for each considered approach.

$$MAE = \frac{1}{N} \sum_{i=1}^N \|y - \hat{y}\| \quad (4)$$

The MSE represents the mean square of the difference between actual and predicted values in the data set, as shown in Equation 5.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2 \quad (5)$$

After each calculation, we observed an improvement in the values of MSE and MAE for the adapted version, in comparison with the original approach (Neto and Silva, 2018). The original approach error percentage was around 21% for the MAE and 7% for the MSE, while in the adapted approach, the MAE value reduced to 8%, and MSE reduced to 1%.

4.3 Data Selection and Extraction

To choose the databases to be used, we considered both open community projects and communities dedicated to women, in order to compare female participation in these two contexts. The open communities were selected through a preliminary survey, considering those which belonged to the same areas as the chosen women dedicated communities.

Data was extracted in January 2022 through a Python application, using the GitHub Rest API⁶.

For the data extraction stage, we used filters based on the work of Neto et al. (2021), divided into two categories: Project Filters (PFs) and Issues Filters (IFs). To select the projects of interest, five Project Filters were defined:

- **PF1.** Project has at least 5 members;
- **PF2.** Project has at least 5 commits;
- **PF3.** Project has at least 5 open issues;
- **PF4.** Project has at least 5 closed issues;
- **PF5.** Project has been created at least 6 months before.

The selection of issues applied the following Issue Filters:

- **IF1.** Issue has at least 5 comments;
- **IF2.** Issue comments contain text besides code snippets;
- **IF3.** Issue has been opened at least one week before;
- **IF4.** Issue has not been closed and reopened.

We emphasize that the PF4 and IF4 filters were defined to ensure that the issues evaluated in the work had their final status assigned as closed, so we could guarantee throughout the data extraction and research that no more comments were inserted. The PF5 filter was chosen in order to favor projects that were minimally operational and performing relevant tasks to the issue tracking system, along with the factors presented in PF2, PF3 and PF4. The IF2 filter was elaborated based on the fact that the comment evaluation is made from their text, i.e., code snippets are not included in the evaluation of the thematic relevance, making comments with only code snippets become irrelevant. Furthermore, it is important to highlight that IF2 takes into account all the comments of a given issue, so that if there is any comment containing only code snippets in an issue, that issue is disregarded.

In the end, data was stored in a CSV file, and later analyzed using Jupyter Notebook.

5 Results and Discussion

After extracting data with the application of the defined filters, we obtained 9151 comments from 1275 issues, present in 28 repositories of 10 communities, 5 of which are open communities and 5 dedicated to women. Table 3 summarizes the data obtained for each community, including the number of repositories, issues, and comments. We emphasize that the first five communities are dedicated to women.

For each of the communities, we calculated the number of female and male collaborators and the Blau index, in order to know their respective diversity factor. The results are presented in Table 4, where it is possible to observe the disparity in terms of female representativeness in the open communities and, consequently, the low index of gender diversity.

As expected, the collected data show low diversity indexes in open communities and high indexes in dedicated communities, since the highest index of 0.5, indicating balance between categories, belongs to RailsGirls. It is also important to highlight that in the PyLadies community, although the Blau index equals 0.0, indicating a lack of diversity, this is

⁶<https://docs.github.com/pt/rest>

Table 3. Overview of data extracted from communities.

Community	Repositories	Issues	Comments
RailsGirls	2	5	148
PyLadies	4	41	360
WomenWhoCode	3	22	168
WoMozBrasil	1	5	26
DjangoGirls	2	145	1014
Django	6	239	1985
Ruby on Rails	4	426	3079
Mozilla	2	81	423
Flutter	1	214	1333
Python	3	97	644

Table 4. Blau index and gender presence by community.

Community	Women	Men	Blau Index
RailsGirls	56	54	0.50
PyLadies	14	0	0.00
WomenWhoCode	3	49	0.11
WoMozBrasil	4	2	0.44
DjangoGirls	9	13	0.48
Django	4	58	0.12
Ruby on Rails	2	66	0.06
Mozilla	14	212	0.12
Flutter	2	82	0.05
Python	9	105	0.15

because the community is closed to male participation, with the aim of creating a safe environment for women to develop both technical and theoretical skills. When pairing Table 3 and Table 4, communities like Ruby on Rails have a high number of posted comments, however, their level of diversity is low. It is also important to note that after an analysis on the GitHub platform, analyzing the first and last names of community members, it was possible to identify that many women who were present in the Ruby on Rails community were also present in the community of the same theme, RailsGirls. This may indicate that participation in the open community was influenced by the empowerment and confidence acquired by participating in the dedicated community.

The other results obtained with the application of metrics CM01 to CM05 focused on the research questions addressed in this study are described in the next subsections.

5.1 RQ1. What is the difference in the participation of men and women in terms of issues and associated comments

First assessing the issues, we identified the gender of all the developers through the gender predicting tool, without any manual intervention. In total, we extracted 1275 issues, out of which 1071 (84%) were reported by men, and 204 (16 %) by women, as shown in Table 5.

Table 5. Comments by community and by gender.

Community	Issues		Comments	
	Women	Men	Women	Men
Dedicated	102	116	741	1166
Open	102	955	1395	5872

Considering the dedicated communities, there were 218 reported issues, out of which 102 (48%) were reported by women. In open communities, 1057 issues were reported, out of which only 102 (9%) were reported by women. Such values show that with the increase of women in the communities, the percentage of their participation increases, unlike men, who have high values in both communities.

In the case of comments, although the authors gender was obtained mostly automatically, manual intervention was necessary in some cases, while only one developer remained with their gender undefined. Manual intervention was carried out for users with unidentified gender using their nicknames on the GitHub platform, in search for additional information that could help in the identification, such as user profile picture, description, and linked personal blogs if present. Furthermore, some comments were posted by bots, and thus not considered in Table 5. From this process, out of the 9151 comments evaluated, 6897 were posted by men, 2059 (22%) by women, and 194 by bots.

In the context of dedicated communities, out of the 1923 comments, 741 (39%) were posted by women. And in the case of open communities, 7447 comments were posted, out of which only 1395 (19%) were posted by women, which also shows low participation compared with men in this segment.

The analyzed comments were filtered to verify the profile of novice developers, shown in Table 6, considering a participation time of less than or equal to 2 years. From this selection, 200 comments were verified as being made by beginner developers, where only 50 (25%) of those comments were posted by women. An important piece of information raised is that looking at the selection of beginner female developers, their percentage of posted comments was higher than their overall result, which was around 22%. This brings evidence that, women may start commenting more in general, but over time these comments decrease, despite the fact the expected tendency would be to increase. In addition, when comparing different communities, the number of comments posted by newcomers in dedicated communities has a smaller percentage, about 25%, than in open communities. However, such comments have a tendency to increase over time, reaching about 39%.

Table 6. Comments posted by beginner developers by community and by gender.

Community	Women	Men
Dedicated	8	12
Open	42	138
Data Base	40	150

Finally, participation in issue tracking was analyzed in the context of the COVID-19 pandemic. Figure 2 shows a graph with the number of comments per year, split by the author's gender. It is observed that despite having a smaller contribution than men, there was a growth rate over the years for women. As shown in Table 7, for the years 2018 and 2019, the growth rate was more than doubled. However, in 2020 and 2021 this growth declined, probably due to the pandemic (Araujo et al., 2022).

It is important to note that the growth rate between 2021 and 2022 was not added to Table 7, since data was extracted

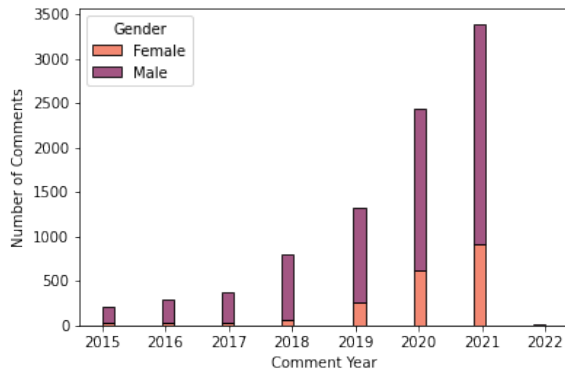


Figure 2. Number of comments posted over the years by gender.

Table 7. Number of comments growth rate over the years.

Years	Women	Men
2015-2016	0.83	1.53
2016-2017	1.50	1.24
2017-2018	1.94	2.16
2018-2019	3.64	1.47
2019-2020	2.41	1.69
2020-2021	1.48	1.36

in January 2022.

5.2 RQ2. Is there a difference between the relevance of comments posted by men and women?

After calculating the thematic relevance for each analyzed comment, the results are similar on average, being 0.03680 for men and 0.03394 for women. As shown in the graph in Figures 3 and 4, it is observed that the data interval changes between genders according to the context of the segment under analysis. In dedicated communities (Figure 3), unlike open ones (Figure 4), women achieve higher relevance values than men.

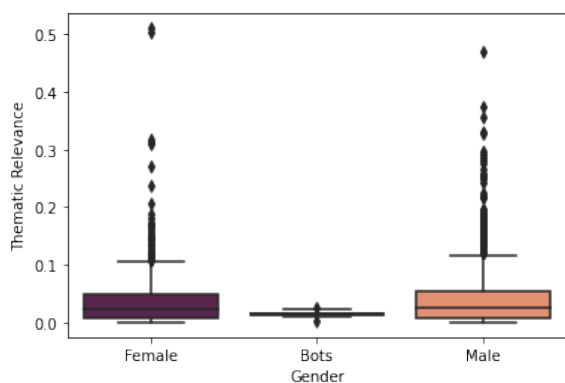


Figure 3. Thematic relevance in dedicated communities.

Results show that although women have a smaller amount of comments posted, as shown in Section 5.1, regardless of the community, open or dedicated, their comments were equally relevant to the discussion around the issues.

Furthermore, looking at beginner developers, despite all the challenges presented by Steinmacher et al. (2012), as we can see in Figure 5, newcomers scored relevance values ranging from 0.0 to 0.5. Another important point to note is

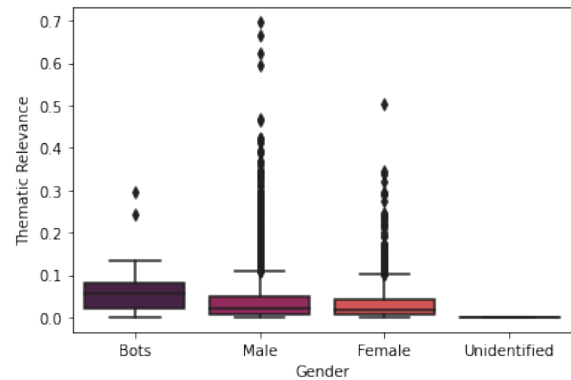


Figure 4. Thematic relevance in open communities.

that comments with relevance in the 0.5 range were posted by women, in both dedicated and open communities.

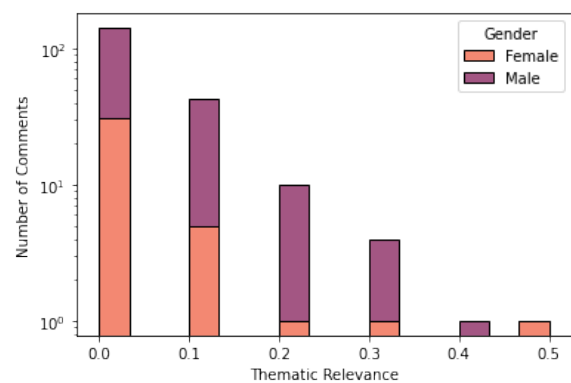


Figure 5. Thematic relevance of beginner developers

5.3 RQ3. Is there a relationship between comment relevance and platform participation time?

To answer this research question, the thematic relevance data and platform participation time were used to construct the graph of Figure 6, in which the columns represent the number of comments made in a certain thematic relevance range, and the black line at the center of each column represents in which years that relevance occurs on average, thus making it possible to observe that up to relevance of about 0.4 the interval is the same, between 6 and 10 years, while higher relevance value ranges show greater variation between years.

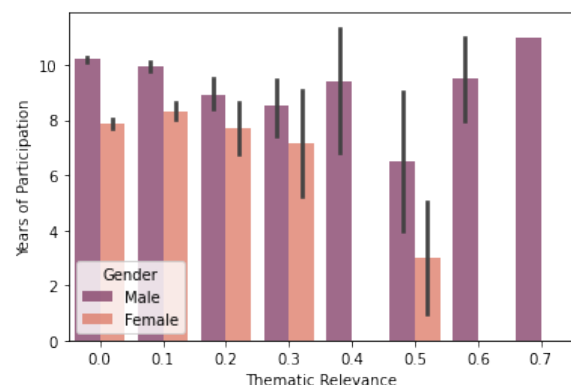


Figure 6. Thematic relevance and years of participation in the platform.

To better evaluate this observation, the Pearson correlation (Paranhos et al., 2014) was applied. Pearson's coefficient ranges from -1 to 1, indicating an increasing trend when positive and close to 1. For the analyzed data, the correlation value between the thematic relevance of the comments and the time of participation in the platform was -0.052, indicating thus the lack of correlation between these variables. Therefore, the comments with higher thematic relevance values are not always made by developers with many years of participation in the platform.

In Figure 6, it is also possible to observe that the intervals representing the participation time of women are smaller than those of men. The range of participation for women is ranging from 2 to about 9 years, while that of men is ranging from 4 years to over 10 years. Such values are reflected both in open communities and in dedicated communities, that is, regardless of the context, women display more recent and smaller participation than men.

5.4 RQ4. Is there a relationship between comment relevance and author's reputation?

Data on thematic relevance and reputation of the comment authors are projected on Figure 7. It is observed that all thematic relevance values have variations according to the comment author reputation, indicating that there does not seem to be a relationship between these two variables. Furthermore, the reputation range present at 0.0 relevance is around 10^3 , and the reputation present at 0.5 relevance comes up to approximately 10^2 , as indicated by the black line in the graph.

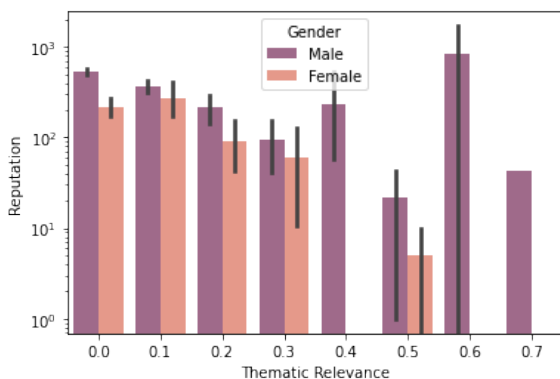


Figure 7. Thematic relevance and reputation.

Again, to confirm this observation, Pearson's coefficient was calculated between the thematic relevance of comments and the reputation of comment authors, resulting in -0.045, once more a low correlation between these two variables. Looking at the selection of beginner developers, in which, as shown in Figure 8, the highest reputation value is around 60, and as shown previously in Figure 5, developers have relatively high thematic relevance of comments despite having little participation time and low reputation values.

Still focusing on Figure 7, we emphasize that women have a smaller reputation interval than men since the range highlighted by the black lines representing women's data reach almost the value of 10^3 in only one case, while the ranges of men's reputation in more than one case almost reach the value of 10^3 , and in others even exceed this threshold. This

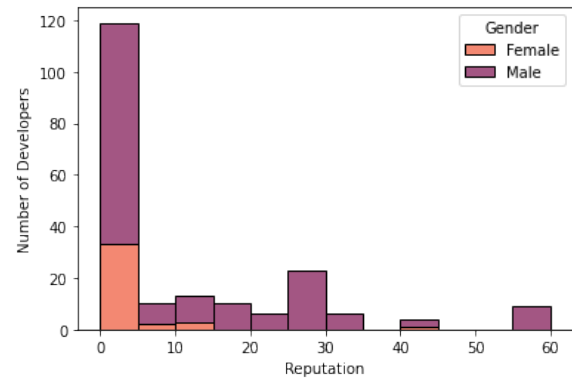


Figure 8. Reputation of beginner developers.

issue can be justified by analyzing Figures 6 and 7 together, in which platform time has an impact on the reputation of developers. To support this analysis, Pearson's coefficient was calculated, obtaining a value of 0.249, which indicates that there is a small positive correlation between these variables. That is, as women have a shorter participation time on the GitHub platform, reaching a maximum of 9 years, they may have a lower reputation since they will consequently make fewer comments, commits, and other activities, which are factors used in the calculation of the reputation metric. It is important to emphasize that both in dedicated communities and in open communities, this scenario is repeated, with the exception that women end up having a similar reputation to men in dedicated communities, where neither reach the value of 10^3 , as shown in Figure 9.

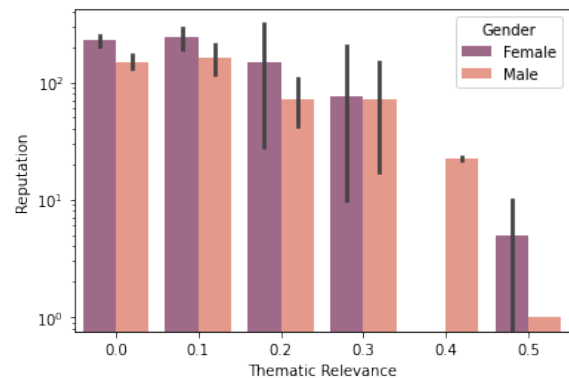


Figure 9. Thematic relevance and reputation in dedicated communities.

6 Threats to Validity

In this work, only the female and male genders were considered for developers. This decision was made to simplify the study, and due to the limitations imposed by the use of the NamSor tool, which can only predict those two genders from a given name.

Regarding the thematic relevance metric, we emphasize that there are still few studies about its usage in the context of issue tracking. In addition, the relatively low values of relevance are influenced by the characteristics of comments posted in issue tracking environments, which may have images, code snippets, content in external links, gratitude messages, among other elements, which are irrelevant to the metric when data is analyzed, or may even not generate enough

similarity with the issue topic.

Finally, regarding the use of the GitScore tool to calculate the CM04 metric, we did not carry out a validation procedure to verify the accuracy of the results obtained, as we did not find a literature work validating this tool.

7 Concluding Remarks and Future Work

In this paper, we present the results of an analysis in issue tracking communication data, noting that, on average, women post comments with similar relevance to those posted by men. The study also reinforced contrasting rates of female participation in open source communities, confirming it in the context of issue tracking, in which women posted only 22% of comments, and authored 16% of issues. In addition, looking at the different evaluated communities, we can see that women have been more active in communities dedicated to them, reaching 48% of the reported issues. However, when we look at the open communities, they have shown low participation and representativeness, with only 9% of the issues reported. An important point about the number of comments posted is the results obtained for the pandemic years, in which even with a smaller number of posted comments, women obtained a high growth rate between 2018 and 2019, before the pandemic. However, from 2020 to 2021, this growth dropped, probably due to the COVID-19 pandemic. In addition, it is important to underline that even for beginner developers, women have a lower number of comments compared to men. Another important point explored by the work is about the reputation and participation time of women in the platform, in which they are usually in an interval of 4 to 10 years, while men have their interval ranging from months to more than 10 years.

As future work, considering technical issues about data analysis, we recommend crossing thematic relevance data with other indicators of participation in software development, such as the number of commits and accepted pull requests. It is possible to assess whether female participation in the scope of communication in projects causes any impact in terms of time taken to solve or close issues. To complement the quantitative analysis of the data, interviews can be carried out with women who participate in both types of communities, to know their perceptions regarding the reasons that lead to the observed low participation rates. Regarding the gender guessing, we suggest the application of other tools such as Genderize, Gender Predict (Zolduoarrati and Licorish, 2021) and genderComputer tool (Catolino et al., 2019) in order to deal with NamSor possible limitations with gender-neutral names. Finally, new works could explore in more depth our initial analysis on women participation during and after the COVID-19 pandemic.

<https://github.com/stardotwav/AnaliseGeneroGitHub> in CSV format, along with the Jupyter Notebook files used to perform the presented analysis.

Acknowledgements

We thank the MinasCoders⁷ program from UFV, Florestal campus, for promoting research involving gender topics in computing.

References

- Araujo, A., Holanda, M., Castanho, C., Koike, C., Oliveira, R., Canedo, E., and Moro, M. (2022). Pandemia de covid-19 tem gênero. In *Anais do XVI Women in Information Technology*, pages 110–121, Porto Alegre, RS, Brasil. SBC.
- Azevedo, B. F. T. (2011). *Minerafórum : um recurso de apoio para análise qualitativa em fóruns de discussão*. Tese de doutorado em informática na computação, Universidade Federal do Rio Grande do Sul.
- Batista, E., e Silva, G. B., and Silva, T. (2022). Diversidade de gênero em projetos open source: um estudo da relevância dos comentários postados em issues do github. In *Anais do XVI Women in Information Technology*, pages 197–202, Porto Alegre, RS, Brasil. SBC.
- Bertram, D., Volda, A., Greenberg, S., and Walker, R. (2010). Communication, collaboration, and bugs: The social nature of issue tracking in small, collocated teams. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW '10*, pages 291–300, New York, NY, USA. ACM.
- Blau, P. (1977). *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, volume 7. Free Press.
- Brandão, M. S., Godinho-Filho, M., Azzolini Junior, W., Battissacco, B. C., and Astorino Marçola, J. (2022). Melhoria da categorização de produtos a partir do uso de algoritmos de aprendizado de máquina e medidas de similaridade. *Revista Produção Online*, 21(4):2093–2124.
- Canedo, E., Tives, H., Bogo, M., Fagundes, F., and Siqueira de Cerqueira, J. (2019). Barriers faced by women in software development projects. *Information*, 10:309.
- Catolino, G., Palomba, F., Tamburri, D. A., Serebrenik, A., and Ferrucci, F. (2019). Gender diversity and women in software teams: How do they affect community smells? In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, pages 11–20, Montreal, QC, Canada. IEEE.
- Izquierdo, D., Huesman, N., Serebrenik, A., and Robles, G. (2019). Openstack gender diversity report. *IEEE Software*, 36:28–33.
- Machado, C. J. R., Maciel, A. M. A., Rodrigues, R. L., and Menezes, R. (2019). An approach for thematic relevance analysis applied to textual contributions in discussion forums. *International Journal of Distance Education Technologies*, 17:37–51.
- Medeiros, D. C., de Queiroz, J. E. R., and Araújo, J. M. F. R. (2014). Análise de funções de similaridade para verificação do conteúdo de mensagens em fóruns de discussão. In *Brazilian Symposium on Computers in Education (Sim-pósio Brasileiro de Informática na Educação - SBIE)*, number 1, page 144, Dourados, Mato Grosso do Sul, Brazil. SBC.

⁷<https://minascoders.caf.ufv.br/>

- Neto, L., Silva, G., and Comarela, G. (2021). Estimativa do tempo de resolução de issues no github usando atributos textuais e temporais. In *Brazilian Symposium on Software Engineering*, page 253–262, New York, NY, USA. Association for Computing Machinery.
- Neto, L. E. C. and Silva, G. B. e. (2018). Colminer: A tool to support communications management in an issue tracking environment. In *Proceedings of the XIV Brazilian Symposium on Information Systems*, SBSI'18, New York, NY, USA. Association for Computing Machinery.
- Noei, E. and Lyons, K. (2022). A study of gender in user reviews on the google play store. *Empirical Softw. Engg.*, 27(2).
- Ortu, M., Hall, T., Marchesi, M., Tonelli, R., Bowes, D., and Destefanis, G. (2018). Mining communication patterns in software development: A github analysis. In *Proceedings of the 14th International Conference on Predictive Models and Data Analytics in Software Engineering*, PROMISE'18, page 70–79, New York, NY, USA. Association for Computing Machinery.
- Outão, J. C. S. d. and Santos, R. P. d. (2022). How does diversity manifest itself in software ecosystems? In *XVIII Brazilian Symposium on Information Systems*, SBSI, New York, NY, USA. Association for Computing Machinery.
- Outão, J. and Santos, R. (2022). Investigando fatores da diversidade de gênero nos ecossistemas de software. In *Anais Estendidos do XVIII Simpósio Brasileiro de Sistemas de Informação*, pages 53–58, Porto Alegre, RS, Brasil. SBC.
- Paranhos, R., Figueiredo Filho, D. B., Rocha, E. C. d., Silva Júnior, J. A. d., Neves, J. A. B., and Santos, M. L. W. D. (2014). Desvendando os mistérios do coeficiente de correlação de pearson: o retorno. *Leviathan (São Paulo)*, (8):66–95.
- Qiu, H. S., Li, Y. L., Padala, S., Sarma, A., and Vasilescu, B. (2019). The signals that potential contributors look for when choosing open-source projects. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW).
- Rodriguez, G., Nadri, R., and Nagappan, M. (2021). Perceived diversity in software engineering: a systematic literature review. *Empirical Software Engineering*, 26.
- Saadat, S., Newton, O. B., Sukthankar, G., and Fiore, S. M. (2020). Analyzing the productivity of github teams based on formation phase activity. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 169–176, Melbourne, Australia. IEEE.
- Singh, V. (2019). Women-only spaces of open source. In *2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE)*, pages 17–20, Montreal, QC, Canada. IEEE.
- Steinmacher, I., Wiese, I. S., Chaves, A. P., and Gerosa, M. A. (2012). Newcomers withdrawal in open source software projects: Analysis of hadoop common project. In *2012 Brazilian Symposium on Collaborative Systems*, pages 65–74.
- Vasilescu, B., Serebrenik, A., and Filkov, V. (2015). A data set for social diversity studies of github teams. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*, pages 514–517, Florence, Italy. IEEE.
- Vedres, B. and Vasarhelyi, O. (2019). Gendered behavior as a disadvantage in open source software development. *EPJ Data Science*, 8.
- Zacchiroli, S. (2021). Gender differences in public code contributions: A 50-year perspective. *IEEE Software*, 38(2):45–50.
- Zolduoarrati, E. and Licorish, S. A. (2021). On the value of encouraging gender tolerance and inclusiveness in software engineering communities. *Information and Software Technology*, 139:106667.