Gender Diversity and Community Smells: A Double-Replication Study on Brazilian Software Teams

Camila Sarmento*, Tiago Massoni*, Alexander Serebrenik†, Gemma Catolino‡, Damian Tamburri‡, Fabio Palomba§
*Federal University of Campina Grande, Campina Grande, Brazil
†Eindhoven University of Technology, Eindhoven, The Netherlands
‡Jheronimus Academy of Data Science, 's Hertogenbosch, The Netherlands
§University of Salerno, Salerno, Italy

***Abstract*—Social debts in software teams are gaining increasing attention from the research community due to their potential adverse effects on software quality. For instance, community smells are indicators of sub-optimal organizational structures and may well lead to the emergence of social debt. Previous studies analyzed which factors influence the emergence/mitigation of such smells. In particular, studies by Catolino et al. showed how factors related to team composition, particularly gender diversity, correlated to the mitigation of community smells. However, a confirmation survey on 60 practitioners suggested that these results were not aligned with the experts' perceptions. In a separate survey, Catolino et al. collected the most common team refactoring strategies for those community smells.**

In this work we replicate two studies by those authors, focusing on the Brazilian software teams; culture-specific expectations on the behavior of people of different genders might have affected the perception of the importance of gender diversity and refactoring strategies when mitigating community smells. We translated the survey instrument used by Catolino et al. to Brazilian Portuguese and recruited 184 Brazilian developers. Results did not show significant differences from the original study; indeed, participants perceived gender diversity as less valuable to mitigate community smells than such factors like experience or team size. Additionally, we performed a qualitative analysis of an open question within the questionnaire for the refactoring strategies. Brazilian developers agree with the original studies for most smells, mainly promoting restructuring communities, creating a communication plan and mentoring.

We believe these results provide further evidence on the problem and its implications when managing software teams, avoiding technical debt and maintenance issues due to team communication and coordination problems.

***Index Terms*—Gender Diversity, Development Teams, Community Smells, Social Debt**

I. Introduction

Diversity is considered as an attribute that distinguishes people in terms of demographic attributes, gender, function, or subjective features (e.g., personality) [1]. From a software development perspective, *gender diversity* is considered as an effective element for creating effectiveness in team performance [2]. For this reason, the research community studied the impact of diversity on software teams [3], [4], [5]. For instance, Santos et al. [6] performed a case study, showing the advantages of gender diversity on teamwork, organizational climate, and critical thinking. These factors need to be carefully monitored in order to avoid the emergence of *social debt* [7], [8], [9], i.e., the appearance of non-cohesive development communities—having communication or coordination issues—that possibly lead to maintenance issues and technical debt in the source code. Social debt has been formalized in terms of *community smells* [10], [11], i.e., a set of observed socio-technical patterns likely leading to the emergence of social debt.

In the context of gender diversity and team composition, Catolino et al. [12] showed how diversity correlated with the mitigation of four different community smells, i.e., Lone Wolf, Organizational Silo, Radio Silence, and Black Cloud. Afterward, they tried to confirm the results through a confirmation survey with 60 practitioners [13]. Findings how gender diversity seems perceived as unimportant when mitigating the presence of community smells compared to other factors, e.g., team size. Later, an additional study [14] surveyed 76 software practitioners on (i) the prominence of the aforementioned four community smells in their contexts and (ii) the methods they adopted to "refactor" them, suggesting that those smells often manifest themselves in software projects, and teams take specific actions to deal with them.

In this work, we conduct a survey with Brazilian developers, replicating questions from the latter two surveys by Catolino et al. [13], [14]. Regarding cultural aspects, several studies found different results regarding organizational variables related to culture [15], [16], [17]. However, little is known about the Brazilian software development community and the impact of culture on gender diversity.

Kohl and Prikladnicki [18] quote national statistical data indicating that 20% of people employed in Brazilian IT are women, and that for the undergraduate courses in Computer Science this drops to 15%. The authors conducted a study on the perceptions of diversity in Brazilian agile software development teams and reported mixed perception of the influence of diversity on the collaboration within a team: some respondents see diversity as beneficial, others—as detrimental. With respect to gender, respondents reported sexism and risk

of sexual tension.

From a sample of 184 Brazilian developers, answering a two-part questionnaire, we first gathered evidence on the perceived importance of gender diversity to mitigate the same selection of four community smells, comparing results with Catolino et al. [13]. Our findings do not show statistical difference between the two samples; indeed, similarly to the original study, Brazilian practitioners perceive gender diversity as less valuable than other factors, i.e., team size and experience.

Second, inspired by the most recent work by Catolino et al. [14], we analyzed open answers of the participants by extracting refactoring strategies for mitigating/solving those community smells. Results show that restructuring the community is the most cited strategy for Organizational Silo and Radio Silence smells. For Black Cloud and Lone Wolf, the preferred alternatives are creating a communication plan and carrying out Mentoring, respectively.

To the best of our knowledge, no other study investigated the relationship between gender diversity and community smells in developing countries. As different cultures induce different expectations and norms for behaviour of people of different genders, our study can confirm, refine or refute previous findings suggesting future research directions.

The rest of the paper is organized as follows: Section II describes state of the art, focusing the attention on the paper replicated. Section III introduces the study design, and the methodology followed to the both replications. Section IV presents the results achieved, discussion and provides remarks and threats to validity. Finally, Section V concludes the paper.

II. Related Work

The effect of gender diversity on software development teams has been gaining more attention from professionals and researchers. Gómez et al. [19] performed a controlled experiment on the productivity of pair programming gender combinations and observed that the productivity rates are similar across established gender pairs. Blincoe et al. [20] noted the positive impact on mood in gender-diverse software development teams.

Regarding the contribution and characterization of women in the OSS community, Wurzelová et al. [21] investigated the potential in the female population contributing and not contributing to OSS, finding no significant differences.

Still exploring development teams and gender diversity participation, Gilal et al. [22] developed a rule-based model for the composition of the software development team, considering the role of the team leader with personality types and gender classification. They concluded that the roles of the software development team vary by gender type.

Regarding community smells, several pieces of research have been carried out. Some studies focus on detecting community smells ([23], [24]), and others evaluate their different types of impacts ([7], [25], [26]). Stefano et al. [27] analyzed the organizational structure patterns of software development teams and their relationship with community smells. Palomba et al. [28] studied the prediction of the emergence of community smells using socio-technical metrics.

Catolino et al.[29] analyzed a set of 40 socio-technical factors (e.g., turnover and communicability) and how they impact four types of community smells. The results suggested communicability as a risk-reducing factor for these smells. Research carried out by Catolino et al. [13] on gender diversity as a reducing factor for community smells obtained results that a team with diverse gender factors seems perceived as unimportant when mitigating the presence of community smells compared to other factors. Another investigation carried out by Catolino et al. [14] was an empirical study that elicited refactoring operations applied by professionals based on the presence of four community smells. Thus, we selected these two papers as an object of replication study to prove the results and consider cultural aspects of Brazilian development teams. In addition to finding new insights, we aim at eliciting guidelines and best practices to improve team collaboration and communication.

III. Study Design

Replication studies are increasingly important for professionals in industry and academia, as they improve the understanding of the processes involved in Software Engineering [30]. Against this background, we report the research questions, objectives, and methodology followed to conduct our double-replication study in this section. All study material, including questionnaires, answer spreadsheets and statistics scripts, can be found in our replication package [31].

To better delineate the methodological process, we divided the methodology into steps: (i) the sampling strategy, where we present the guidelines that we follow and the approach taken to collect data; (ii) data collection, in which we present the characterization of the data collected and the description of the research instrument and (iii) the analysis methods, how we conduct our analysis process to carry out the replication.

A. Research Questions and Goals

The purpose of this research is twofold. Our *first motivation* is analyzing the perceptions of Brazilian practitioners when evaluating gender diversity as a factor for mitigating community smells—replicating the study of Catolino et al. [13]. Our *second motivation* is to propose refactored strategies based on the replication of the study of Catolino et al. [14] with the aim of findings new insights, thus eliciting guidelines and best practices to improve collaboration and communication of teams. To achieve our goal, we pose the following research questions:

RQ1. *What is the perception of software developers concerning the relevance of gender diversity to address the occurrence of four Community Smells (Organizational Silo, Black Cloud, Lone Wolf and Radio Silence)?*

RQ2. *What are the main actions taken by Brazilian software development teams to mitigate those Community Smells?*

To answer these research questions, we conduct a survey combining Likert scale questions (**RQ1**) and open questions

(**RQ2**). In **RQ2** we opt for open questions and a subsequent coding rather than confronting the respondents with the codebook of Catolino et al. [14] as not to drive the respondents.

B. Community Smells

We focus on the same community smells considered in the original studies [13], [14]: (*i*) *Organizational Silo* occurs when the two siloed subgroups of the team do not communicate, except through one or two of their respective members; (*ii*) *Black Cloud* occurs when team members experience information overload due to lack of structured communication; (*iii*) *Lone Wolf* occurs when an unsanctioned or defiant contributor carry out their work with little consideration of their peers, their decisions and communication; and (*iv*) *Radio Silence* when one member interposes herself into every formal interaction across two or more sub-communities with little or no flexibility to introduce other parallel channels.

C. Sampling

As a methodological sampling strategy, we followed the Guidelines for Conducting Software Engineering Research proposed by Stol and Fitzgerald [32] in a field context, thus addressing the study's purpose and observing our variables of interest (Gender, Refactoring strategies) and control variables (Experience and Team Size).

Sampling was done in a non-probabilistic way, following a mixed approach: *Convenience*, selecting participants based on availability, and *Referral-chain*, also called snowball sampling, with the development teams, therefore, worked by finding a few survey respondents in the population, inviting to answer the survey, and asking them to identify other members of the population [33]. Consequently, professionals were recruited in two stages. The first step consisted of sending the survey to close contacts; the second, asking the volunteer respondents to disclose the invitation to their colleagues and report the number of invitations sent to us. We advertised the survey through WhatsApp, Facebook, LinkedIn, email and Slack.

Some contacts requested to be reminded at other times, and others took two attempts to complete the survey. The highest rate of responses occurred in the afternoon; specifically, between 11 am and 2 pm. We obtained some responses early morning from the links sent after 9 pm.

In this way, we had an average time of eight minutes of survey response and achieved a completion rate of 66%. This completion rate was informed by the tool used for data collection (survey monkey[1]) which provided statistics about participants who started their contribution but, at some point, gave up on completing their answers and did not submit.

Using the approach mentioned above, we approached 194 participants and obtained 67 responses. In the snowballing process, based on the volunteers reporting the number of people to whom they sent the survey, we estimate the number of survey invitations as ca. 1.752 participants (some participants sent to groups on WhatsApp and Slack); in this way, we obtained 117 responses. In total, 1,946 professionals were approached, and we have received 184 responses.

The Research Ethics Committee has approved the study under a substantiated opinion and Certificate of Presentation and Ethical Appreciation (CAAE), which was created to defend the interests of research participants in their integrity and dignity and to contribute to the development of research within Brazilian ethical standards.

D. Data Collection

We adopted the instrument consisting of four vignettes [13], [14] corresponding to the four community smells:

- *Scenario 1 - Organizational Silo*: "Suppose your development team is working on the definition of a web-based application for the scheduling of resources. During the development, you recognize the existence of independent sub-teams that do not communicate with each other except through one or two of their respective members."
- *Scenario 2 - Black Cloud*: "Suppose your development team is working on the definition of a web-based application for the scheduling of resources. During the development, you recognize that the community members suffer of an information overload due to lack of structured communication (e.g., communications among team members are not performed over official channels)."
- *Scenario 3 - Lone Wolf*: "Suppose your development team is working on the definition of a web-based application for the scheduling of resources. During the development, you recognize the presence of a individual who carries out her work independently from the decisions taken by the community."
- *Scenario 4 - Radio Silence*: "Suppose your development team is working on the definition of a web-based application for the scheduling of resources. During the development, you recognize that one developer, belonging to a sub-team A, interposes herself into every formal communications across two or more sub-teams, with little flexibility to introduce other parallel communication channels."

Similarly to Catolino et al. [13], [14] each vignette was followed by Likert-scale questions (from "Not at all important" to "Very Important") for gender diversity, developer experience and team size, and open questions about the motivation for the classification and suggestion of additional factors that might help mitigate the problem.

Once the vignettes had been translated into Portuguese, we applied four pilot tests to verify that the translation of the questionnaire was interpreted as expected. After administering the pilot tests, we noted the need to change some terms from the translation to the reality of Brazil in the demographics section, e.g., from *"Operations Team Member"* to *"Support Team Member"*. We added the question: *"In which Brazilian state do you currently work in?"* to ensure an adequate spread of the respondents over different regions of Brazil. Table I provides an overview of the final survey instrument. The survey was applied in Brazil between February and May 2021.

[1]https://pt.surveymonkey.com/

TABLE I
SUBSET OF THE SURVEY INSTRUMENT.

Analysis	Comments	Description	Type
Scenarios	Please rate the importance of the following aspects for mitigating this smell	Presence of Diverse Gender	Closed
		Experience	Closed
		Team Size	Closed
		Motivate the previous answers, adding if you believe that other factors might mitigate these issues	Open
Demographics	A bit about yourself	Gender	Closed
		Role that best describes your current job	Closed
	With respect to your peers, how do you rate your experience in...	Team Management	Closed
		Software Development	Closed
	Team and Company info	Company Size	Closed
		Team Size	Closed

E. Analysis

For each scenario, we present an open and non-mandatory question to respondents (see Table I): "Motivate the previous answers, adding if you believe that other factors might mitigate these issues." Then, we obtained a total of **264** written sentences.

1) ***RQ1 analysis****:* In the demographic analysis, data were collected from closed research questions, and only the question about the respondent's gender remained open and non-mandatory. Regarding the data collected from the closed questions (quantitative data), we first check whether the distribution of the response data is normal and then use the appropriate statistical tests.

Normality was checked based on the Shapiro-Wilk Test [34] with α equal to 0.05. After verifying the non-normality of the data, we used the Kruskal-Wallis [35] statistical test by sampling groups to observe their differences. In addition, we compare our results with the original article [13]. All inferential tests were previously set at 0.05. As for the effect size, whose value was calculated from Cohen's d [36].

2) ***RQ2 analysis****:* Based on the survey's open responses, we accessed qualitative data about the motivation of the participants regarding the closed responses attributed to the smell scenarios and about suggestions for actions that can mitigate community smells (see Table I). The answers to open questions were based on the perspective of the code book proposed by Catolino et al. [14] which provides specific refactoring practices to deal with each of the community smells considered in this investigation.

In addition to the open coding method applied in the original study, we employed the Delphi-inspired approach [37] for conflict resolution, following the approach used in Tamburri et al. [38]. Therefore, our analysis remains comparable to the original results, whose coding process was consisted of three primary cycles (see Figure 1): paired coding; conflict moderation and Delphi approach.

The first cycle, *paired coding*, has been carried out by the first four authors. Two hundred sixty-four sentences were distributed into six groups corresponding to six pairs of raters. Each rater performed coding independently according to the referential codebook [14]. Subsequently, we found 79 disagreements.

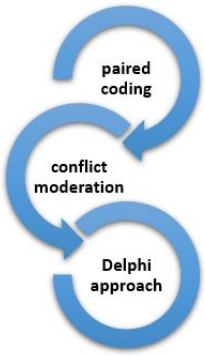

Fig. 1. Coding process.

Next, we invited two other authors to moderate the conflicts. This reduced the number of disagreements to 23. Hence, before proceeding with the Delphi process, we perform a dialogic inter-subjectivity to align the respective definitions. The procedure took place at a meeting and, later, the raters added comments to the codes, justifying the reason for the code applied.

Based on this principle, 8% of the codes remained in conflict, so we followed the approach used by Tamburri et al. [38], in carrying out the *Delphi analysis* process. Our orchestration occurred as follows: we performed a voting system between three reviewers for each code group—only those that had conflicts—, and the most voted option among the three reviewers was selected.

IV. RESULTS AND DISCUSSION

In this section, we present and discuss the results obtained from the performed double-replication study, in terms of the proposed research questions (RQs).

A. *RQ1. What is the perception of software developers concerning the relevance of gender diversity to address the occurrence of four Community Smells (Organizational Silo, Black Cloud, Lone Wolf and Radio Silence)?*

To present and discuss our results, we have organized the subsection into two parts: initially, we exhibit our quantitative results and discuss their implications; later, we compare our results with the original paper [13].

1) Quantitative results and implications: After observing the non-normality of the data and the asymmetry degree of the sample, non-parametric tests were applied, along with determining effect sizes, as performed using the IDE Open-Source RStudio[2]. Considering the degree of distribution and flatness of the collected sample, skewness and kurtosis were observed. Scale values from Gender, Experience and Team Size are asymmetric with sharp peaks, except for categorized data for Lone Wolf's team size answers.

By statistically testing the variance between groups and within groups of Community Smells for each scale, the results show *significant differences among Gender, Experience, and Team size*, indicating participants showed distinct preferences of each scale. Based on these results, we made multiple statistical comparisons between the paired groups. (see Table II). As such, we can see that the mean ranks of the pairs are significantly different, i.e., various groups assign distinct importance degrees according to the type of Community Smell. We also noticed that **Gender** is generally considered to be the **least important** factor to mitigate all Community Smells (positive effect sizes against Experience and Team Size).

TABLE II
KRUSKAL-WALLIS TEST RESULTS, PAIRWISE COMPARISON.

Pair	Community Smell			
	OS	**BL**	**LW**	**RS**
Gender and Experience				
p-value	4.89E-12	0	0	0
Effect size	0.16	0.26	0.30	0.22
Gender and Team Size				
p-value	1.411E-6	0	7.69E-09	1.55E-06
Effect size	0.061	0.19	0.13	0.097
Experience and Team Size				
p-value	9.955E-6	6.609E-3	1.31E-04	2.27E-5
Effect size	0.051	0.017	0.073	0.046
Consider $\alpha < 0.05$				

Analyzing the statistical comparisons by effect size (see Table II) it is possible to notice the difference in the participants' responses for each community smell. Survey respondents inequivocally suggest that Experience and Team Size are most essential to mitigate Organizational Silo.

The effect size of each factor for Black Cloud and Lone Wolf is similar. Participants indicate that the lack of structured communication and solo work are mitigated more efficiently with more experienced people on the team than with the presence of people of different genders.

For Radio Silence, the experience was seen as the factor that primarily influences the mitigation effect, followed by Team Size and Gender. As opposed to other community smells, Team Size appears to be less critical.

These results indicate gender diversity is perceived as less important than experience and team size also by Brazilian developers, agreeing with the results from the first replicated study. The majority of the participants promoted the experience of team members as an essential measure against the presence of community smells, as well as of the size of the team.

2) Comparison with Catolino et al.[13]: Figure 2 (*Organizational Silo*) suggests differences in the results for Gender Diversity and Experience, as it is clear that both the first and the third quartiles of the Brazilian boxplot are higher. However, both studies maintain the relative statistical results for these factors.

Figure 3 (*Black Cloud*) shows that both the original study and in the replication, Experience and Team size are perceived as more important than gender diversity. The boxplots suggest a similar perception of the importance of individual factors.

Figure 4 represents the smell assessments related to *Lone Wolf*. It is possible to notice the same orders of importance attributed to the factors. The original survey participants seemed to have attributed lower importance to gender diversity and team size than their Brazilian counterparts.

Figure 5 (*Radio Silence*) suggests the same statistical results of the three factors. Brazilian respondents seem to attribute higher importance of all three factors studied to mitigation of this community smell.

Although the original and replicated data have distributed opinions, the categorizations of results do not diverge; the original paper concluded that the participants did not perceive the factor of gender diversity as important to mitigate community smell scenarios, whose effect is equally applied in Brazil. However, we can highlight the experience factor of replicated data, classified as the most critical factor to be considered to mitigate the effects of community smells.

B. *RQ2.What are the main actions indicated by Brazilian software development teams to mitigate the community Smells?*

To answer our second research question, corresponding to the second replicated study from Catolino et al. [14], we refactored the strategies by evaluating the open survey responses. In this study, we used MAXQDA Standard for Windows[3] to import the text files to organize and categorize the data for better visualization.

Our first observation from the quotes occurred from the separation of the positive considerations regarding the influence of gender diversity in reducing the number of community smells and contrary comments. Thus, we came across a rate

[2]rstudio.com

[3]https://www.maxqda.com

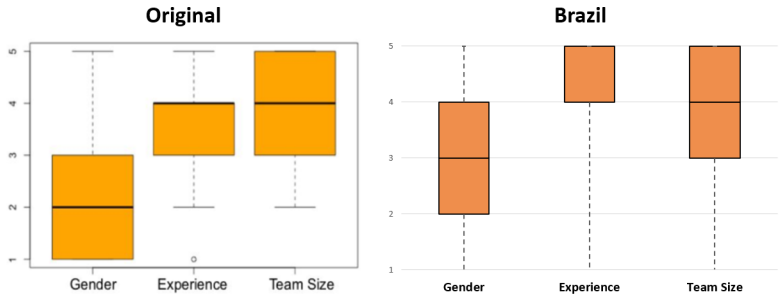

Fig. 2. Organizational Silo Evaluation.

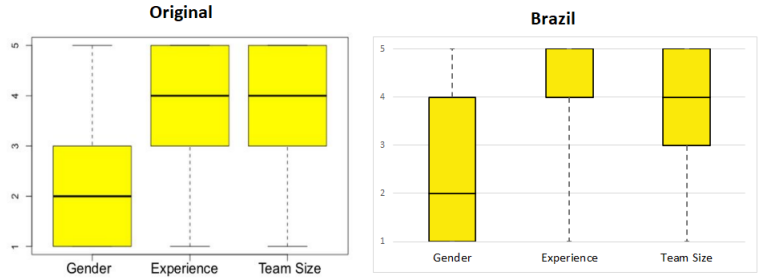

Fig. 3. Black Cloud Evaluation.

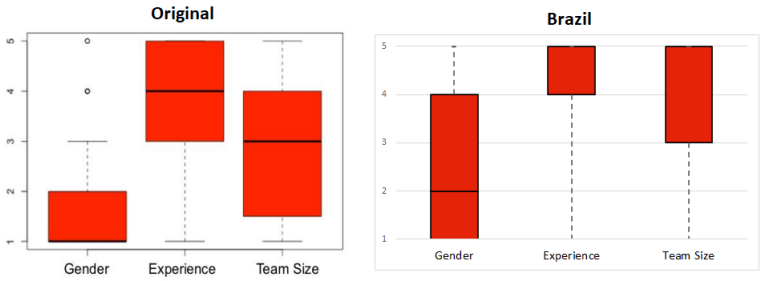

Fig. 4. Lone Wolf Evaluation.

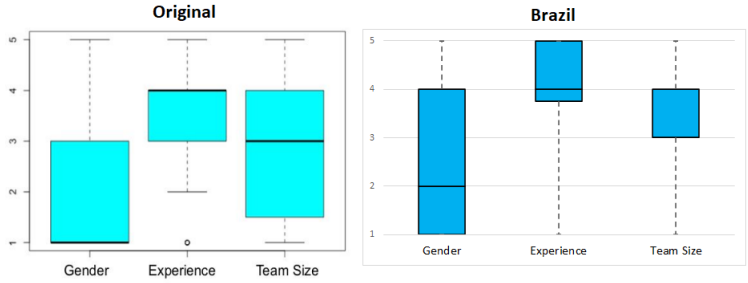

Fig. 5. Radio Silence Evaluation.

of 47.4% of quotes that mentioned the positive influence of gender diversity, while 52.6% stated that this was not an important factor. Given this context, we can see that, despite this difference presented, we cannot say with certainty.

So, for better analysis, we isolated the data according to their respective community smell. Thus, it was possible to note that for the participants, gender diversity is an essential factor only for the scenario that the team has failures in communication, except through one or two respective members, even with the same project; which is associated with Organizational Silo Smell. For the other smells, participants did not consider gender diversity a determining factor for

reducing the community smell problem.

Still, considering the other factors in this research (Experience and Team Size), it is evident that experience is considered as a reducing factor of community smells, whose codes assigned by factor are shown in Table III.

TABLE III
CODE ASSIGNED BY FACTOR.

Factor	Community Smell			
	OS	BL	LW	RS
Diversity	42,9%	4.2%	8.4%	62.5%
Experience	33,3%	62.5%	58.3%	25.0%
Team size	23,8%	33.3%	33.3%	12.5%

In addition to gender diversity, some participants mentioned the importance of other types of diversity, such as P80 *"Diversity of all types is very important for defining project scope and vision"* and P2 *"The greater diversity in a team, The greater the possibility of different perceptions and ideas, which could improve communication and exchange between teams.), etc"*. Still, even citing diversity as a relevant factor, some participants also pondered the importance of experience, as highlighted by P34 *"people of different genders, as well as people with different backgrounds and experiences, add to the team a greater capacity for interaction and engagement"*.

In addition, the Team Size was the factor considered to have the most negligible influence on the reduction of community smells, as mentioned by P77: *"Team size is irrelevant in a scenario like this [...]"* and P53 *"Mixed teams tend to do better. It doesn't matter the number of members"*.

During the exploration and analysis of data from open survey responses, from a refactoring process of replicated codes in the light of the Codebook elucidated by Catolino et al. [14], it was possible to outline strategic actions that can prevent community smells in development teams. It is noteworthy that about the answers that did not present answers to the problem, they were considered as outliers, or rather, *Not relevant*.

1) Organizational Silo: Table IV scoring lists of refactoring strategies for the Organizational Silo. In this case, it was possible to analyze which strategies are identified as the most effective to prevent such communication failure. Thus, we can indicate that, for this smell, the most relevant strategy for prevention is the restructuring of the community, which includes organizing sub-teams heterogeneously, splitting larger teams into smaller ones and swapping team members for every task/story/sprint. This result corroborates with the original study [14]. According to the participants ' opinion, the strategy considered less relevant is the introduction of a social-rewarding mechanism.

For application in the industrial environment, there is the possibility of orchestrating strategies simultaneously. In these cases, the participants strongly recommend restructuring the community and, as a secondary alternative, a Mentoring approach.

TABLE IV
STRATEGIES FOR THE ORGANIZATIONAL SILO SMELL.

Strategy	**#**
Introduce a social-rewarding mechanism	01
Monitoring	06
Create communication plan	07
Cohesion exercising	10
Mentoring	12
Restructure the community	32
Legend:	
# Number of strategies.	

2) Black Cloud: The strategy indicated by the participants that are most effective in mitigating the lack of structured communication is "*Creating a communication plan*" as shown in Table V, which echoes the original study [14]. According to the taxonomy proposed by that paper, it corresponds to creating appropriate communication channels and protocols.

TABLE V
STRATEGIES FOR THE BLACK CLOUD SMELL.

Strategy	**#**
Create communication plan	27
Restructure the community	17
Introduce a social sanctioning mechanism	08
Legend:	
# Number of strategies.	

On the other hand, with the same result for the Organizational Silo, "*Introduction of a social sanctioning mechanism*" corresponded as the strategy with the lowest effect. Apparently, the introduction of a social sanction mechanism may already exist in the organization. Hence, a lack of structured communication occurs, so it was the least mentioned strategy as a solution.

3) Lone Wolf: When a team member works independently and is isolated from the rest of the team, in the smell considered as Lone Wolf (see Table VI), the strategy "*Mentoring*" was the most indicated by the participants, which was observed in the replicated study [14]. Nevertheless, reflecting on the second strategy, the difference between "*Restructure the community*" and "*Introduce a social sanctioning mechanism*" is very subtle, with only one opinion of difference, which makes it difficult to assert the second strategic indication reliably.

TABLE VI
STRATEGIES FOR THE LONE WOLF SMELL.

Strategy	**#**
Mentoring	16
Restructure the community	11
Introduce a social sanctioning mechanism	10
Monitoring	05
Cohesion Exercising	05
Legend:	
# Number of strategies.	

4) Radio Silence: The most cited strategy to mitigate this smell, according to the participants (see Table VII), is the restructuring of the community. On the other hand, unlike other smells, the least efficient strategy indicated by the participants is *Monitoring*; this result makes sense because when there is a high formality of regular procedures due to structural disorganization, and monitoring the team will not reduce this problem. In the replicated study [14], Cohesion Exercising was considered the most relevant strategy for tackling Radio Silence; similarly, Monitoring was also regarded as of minor importance in that study.

TABLE VII
STRATEGIES FOR THE RADIO SILENCE SMELL.

Strategy	**#**
Restructure the community	14
Create communication plan	07
Mentoring	06
Introduce a social sanctioning mechanism	04
Cohesion Exercising	01
Monitoring	00
Legend:	
# Number of strategies.	

From the various strategies listed, we can see that there is no single strategy for all purposes. For the Organizational Silo and Radio Silence Smells cases, the best option would be to restructure the community. On the other hand, when the team is faced with information overload due to a lack of structured communication (Black Cloud), the best strategy is to create a communication plan. For cases where team members work independently and in isolation from others, Lone Wolf, the ideal would be to execute a mentoring program.

C. Demographics

Each participant was asked to provide their gender and current job. Also, they provided an estimate for experience in both team management and software development, in addition to team and company size.

As for the role that best describes the participant's current job (see Table VIII), the vast majority have been identified as *"Development Team Member"* (*96* participants). In the case of the *"others"* role, the respondent had the option to fill in their role manually. In these cases, an accumulation of functions emerges, e.g., Teacher/Technical Leader, or positions within the framework of *"Development Team Member"*, such as Machine Learning Engineer and Data Scientist. Such circumstance occurs due to the lack of standardization in Brazil.

Following the original article, which collected experience as a Likert scale from basic to considerable, we group the responses about management experiences into a bar chart (Figure 6). The sample is diverse in terms of experience in team management, with a slight predominance of professionals with *Considerable* (28%) and *More than good* (34%) experience in team management. A similar result was observed for development experience, in which most participants consider them as considerably experient (40%).

TABLE VIII
ROLES OF PARTICIPANTS

Role	**#**	**%**
Development Team Member	96	51.3%
Project Manager	14	7.5%
Member of the Quality/Testing team	12	6.4%
Software architect	12	6.4%
Coach / Scrum Master	7	3.7%
Product owner	7	3.7%
Support Team Member	5	2.7%
Other	31	18.2%
Legend:		
# Number of functions		
% Percentage of functions		

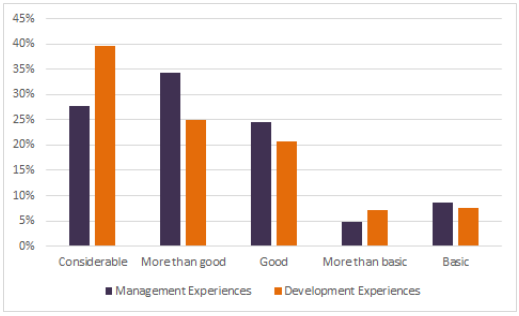

Fig. 6. Management and Development Experiences

About the participants' *gender*, the sample was composed of *42* (23%) women and *101* (55%) Men. We grouped the gender info with company size and team size in Figures 7 and 8. This data is entirely consistent with the characteristics of the Brazilian software industry. Women have higher participation in companies between 01-49 collaborators and more than 2000 collaborators. In the representation of team size *versus* gender, the involvement of women is more prevailing in teams between 02-05 and 05-10 collaborators, decreasing for the other team sizes.

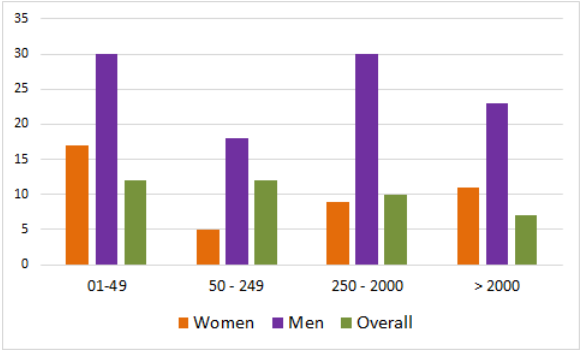

Fig. 7. Company size by gender

Considering the representativeness of the sample, our sample comprised fourteen states (see Table IX). There is a predominance of the sample in certain regions. This is due to the concentration of developers in certain states and the convenience of the sample. The state of Paraíba stands out

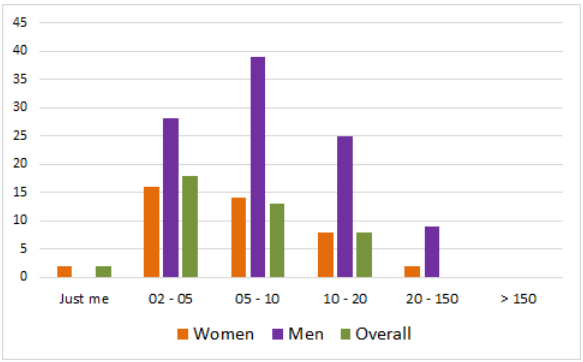

Fig. 8. Gender by Team Size

with *106* respondents, reflecting the state that started the call for participation.

TABLE IX
SAMPLE CHARACTERIZATION BY REGION..

State	#
Paraíba	106
São Paulo	033
Minas Gerais	011
Rio de Janeiro	008
Pernambuco	005
Bahia	004
Rio Grande do Sul	004
Santa Catarina	004
Goiás	002
Paraná	002
Rio Grande do Norte	002
Alagoas	001
Amazonas	001
Mato Grosso do Sul	001
Legend:	
# Number of surveys respondents.	

This convenience approach was carried out in the states of Paraíba, Pernambuco and Minas Gerais. As for the other states represented in the sample, data were collected using the Referral-chain process [39]. It is also noteworthy that the data are not exclusive — there may be data collected from the Referral-chain process in the states where we started the approach for convenience.

D. Threats to Validity

In performing our double-replication study, we reduced some threats to validity, trying to avoid compromising our ability to compare the data. For *internal validation*, we applied the same material provided by the replicated studies. We applied a pilot survey with four participants mainly for validating our translation to Portuguese of the questionnaire. Regarding *external validity*, the study reflected the diversity of development teams in Brazil that had different characteristics, namely: gender, experience levels of managing, experience levels of developing, organization sizes, team sizes and functions. Even so, generalization is not a possible claim.

In presenting the survey, we inform the volunteer participants that the survey is anonymous and that the data collected will be analyzed without revealing the identity of the respondents. As it is a mixed-approach study, the collected data was observed their normality's and applied the respective statistical comparison tests for comparison. Also, we could code the open questions of the survey to carry out our second research question.

There is a predominance of the sample in certain regions. This is due to the concentration of developers in certain states and the convenience of the sample. The state of Paraíba stands out with *106* respondents, reflecting the state that started the call for participation. This convenience approach was carried out in the states of Paraíba, Pernambuco and Minas Gerais. As for the other states represented in the sample, data were collected using the Referral-chain process [39]. It is also noteworthy that the data are not exclusive — there may be data collected from the Referral-chain process in the states where we started the approach for convenience.

To mitigate the threat of *validity to the conclusion* related to qualitative analysis, five experts in the field participated in the coding process under various methods until we reached the conclusions.

V. CONCLUSION

In this paper, we replicate two studies on the relationship of Gender Diversity and Community Smells in software teams [13], [14]. For replicating Catolino et al. [13], we assess whether Brazilian developers also, as in the original study, regard diversity as less valuable to teams than experience or team size when dealing with four community smells (Organizational Silo, Black Cloud, Lone Wolf and Radio Silence). In turn, we replicate a second paper [14], by analyzing qualitatively several open answers provided by the participants on how they would address the hypothetical occurrence of those smells in their teams.

As a result, we collected 184 answers from participants recruited through hub contacts working in several regions of Brazil. In comparison with the original study, we found the same general levels of agreement concerning community smells. Distributions were approximate — the analysis of the effects did not reveal a significant difference. It means that, quantitatively, Brazilian developers also think diversity is not as important as team size and experience to mitigate the effects of community smells in all scenarios.

In a qualitative analysis, participants mentioned diversity is relevant only for the Organizational Silo smell. Regarding the strategies to mitigate smells, for Organizational Silo and Radio Silence, predominantly due to communication issues, the most effective action for the leader to operate would be the restructuring of the community. The solution pointed to the Black Cloud smell was more evident: create a communication plan. But also, we cannot disregard the second opinion of the participants, which was the restructuring of the community, as there was more emphasis on the last solution option (introduce a social sanctioning mechanism) and also because it corresponds to 32.6% of the opinions. For the case of smell Lone Wolf, respondents mainly emphasized Mentoring as a

strategy. For the action of a leader in these cases, we can consider an order of actions that leaders could employ. First, a mentoring with the team member; if the smell is not resolved, the leader could restructure the community or introduce a social sanctioning mechanism, in that order.

One of the original papers [13] triangulates the survey with practitioners with the results of previous evidence on how the presence of women could improve team communication and collaboration. In our replication, the participants also perceive gender diversity as less important than experience or team size to mitigate community smells. However, several participants showed signs of understanding why diversity is an essential measure to address social debt in the open answers. More empirical evidence is needed to make diversity important to teams; our study highlights how strong evidence is for theory assimilation in practice, regardless of country or culture. Additionally, most participants mentioned communication skills are essential to tackle community smells in software teams. The promotion of communication as a primary skill must then be strongly considered in hiring and training.

Other community smells will be considered in our future research efforts, mainly associated with gender diversity. Also, future research is needed to determine this difference and identify if the experience is an essential factor that can mitigate community smells. Software team leadership must have a deeper understanding of the trade-offs among developers' experience, gender diversity and other social-related choices. This replication and the related studies serve as valuable exercises for practitioners to reassess and restructure team practices in general.

ACKNOWLEDGMENT

Gemma and Damian are supported by the European Commission grant no. 825040 (RADON-H2020). Fabio acknowledges the support of the Swiss National Science Foundation through the SNF Project No. PZ00P2_186090 (TED).

REFERENCES

[1] M. Ortu, G. Destefanis, S. Counsell, S. Swift, M. Marchesi, and R. Tonelli, "How diverse is your team? investigating gender and nationality diversity in github teams," PeerJ Preprints, Tech. Rep., 2017.

[2] A. R. Gila, J. Jaafa, M. Omar, and M. Z. Tunio, "Impact of personality and gender diversity on software development teams' performance," in *2014 International Conference on Computer, Communications, and Control Technology (I4CT)*, 2014, pp. 261–265.

[3] N. Pinheiro and K. Gama, "Diversidade e inclusão: cultura e percepção em empresas de tecnologia da informação," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, pp. 1–1, 2020.

[4] K. Blincoe, O. Springer, and M. R. Wrobel, "Perceptions of gender diversity's impact on mood in software development teams," *IEEE Software*, vol. 36, no. 5, pp. 51–56, 2019.

[5] K. Kohl Silveira, S. Musse, I. H. Manssour, R. Vieira, and R. Prikladnicki, "Confidence in programming skills: Gender insights from stackoverflow developers survey," in *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, 2019, pp. 234–235.

[6] A. L. Santos, J. D. Alexandre, and F. Meireles, "A survey about gender diversity in a more gender-balanced it firm," in *2021 4th International Conference of the Portuguese Society for Engineering Education (CISPEE)*, 2021, pp. 1–4.

[7] F. Palomba, D. A. A. Tamburri, F. A. Fontana, R. Oliveto, A. Zaidman, and A. Serebrenik, "Beyond technical aspects: How do community smells influence the intensity of code smells?" *IEEE transactions on software engineering*, 2018.

[8] D. A. Tamburri, P. Kruchten, P. Lago, and H. van Vliet, "What is social debt in software engineering?" in *2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*. IEEE, 2013, pp. 93–96.

[9] D. A. Tamburri, P. Kruchten, P. Lago, and H. Van Vliet, "Social debt in software engineering: insights from industry," *Journal of Internet Services and Applications*, vol. 6, no. 1, pp. 1–17, 2015.

[10] D. A. Tamburri, P. Lago, and H. v. Vliet, "Organizational social structures for software engineering," *ACM Computing Surveys (CSUR)*, vol. 46, no. 1, pp. 1–35, 2013.

[11] D. A. Tamburri, "Software architecture social debt: Managing the incommunicability factor," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 1, pp. 20–37, 2019.

[12] G. Catolino, F. Palomba, D. A. Tamburri, A. Serebrenik, and F. Ferrucci, "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)*. IEEE, 2019, pp. 11–20.

[13] ——, "Gender diversity and community smells: insights from the trenches," *IEEE Software*, vol. 37, no. 1, pp. 10–16, 2019.

[14] ——, "Refactoring community smells in the wild: the practitioner's field manual," in *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Society*, 2020, pp. 25–34.

[15] Z. Aycan, R. Kanungo, M. Mendonca, K. Yu, J. Deller, G. Stahl, and A. Kurshid, "Impact of culture on human resource management practices: A 10-country comparison," *Applied Psychology*, vol. 49, no. 1, pp. 192–221, 2000. [Online]. Available: https://iaap-journals.onlinelibrary.wiley.com/doi/abs/10.1111/1464-0597.00010

[16] I. S. Windiarti, A. Prabowo, M. Qamaruzzaman *et al.*, "Women in informatics engineering career: Perspective from hofstede cultural dimension and dayak tribe's cultural values," in *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE, 2018, pp. 278–282.

[17] B. McSweeney, "Hofstede's model of national cultural differences and their consequences: A triumph of faith-a failure of analysis," *Human relations*, vol. 55, no. 1, pp. 89–118, 2002.

[18] K. Kohl Silveira and R. Prikladnicki, "Perceptions on diversity in brazilian agile software development teams: A survey," in *Int Workshop on Gender Equality in Software Engineering*, 2018, pp. 37–40.

[19] O. S. Gómez, M. Solari, C. J. P. Calvache, and A. C. Ledezma, "A controlled experiment on productivity of pair programming gender combinations: Preliminary results." in *CIbSE*, 2017, pp. 679–692.

[20] K. Blincoe, O. Springer, and M. R. Wrobel, "Perceptions of gender diversity's impact on mood in software development teams," *Ieee Software*, vol. 36, no. 5, pp. 51–56, 2019.

[21] P. Wurzelová, F. Palomba, and A. Bacchelli, "Characterizing women (not) contributing to open-source," in *2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE)*. IEEE, 2019, pp. 5–8.

[22] A. R. Gilal, J. Jaafar, M. Omar, S. Basri, and A. Waqas, "A rule-based model for software development team composition: Team leader role with personality types and gender classification," *Information and Software Technology*, vol. 74, pp. 105–113, 2016.

[23] N. Almarimi, A. Ouni, M. Chouchen, I. Saidani, and M. W. Mkaouer, "On the detection of community smells using genetic programming-based ensemble classifier chain," in *Proceedings of the 15th International Conference on Global Software Engineering*, ser. ICGSE '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 43–54. [Online]. Available: https://doi.org/10.1145/3372787.3390439

[24] N. Almarimi, A. Ouni, and M. W. Mkaouer, "Learning to detect community smells in open source software projects," *Knowledge-Based Systems*, vol. 204, p. 106201, 2020.

[25] D. Tamburri, F. Palomba, and R. Kazman, "Exploring community smells in open-source: An automated approach," *IEEE Transactions on Software Engineering*, no. 01, pp. 1–1, feb 2019.

[26] B. Eken, F. Palma, B. Ayşe, and T. Ayşe, "An empirical study on the effect of community smells on bug prediction," *Software Quality Journal*, vol. 29, no. 1, pp. 159–194, 2021.

[27] M. De Stefano, F. Pecorelli, D. A. Tamburri, F. Palomba, and A. De Lucia, "Splicing community patterns and smells: A preliminary study," in *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops*, 2020, pp. 703–710.

[28] F. Palomba and D. A. Tamburri, "Predicting the emergence of community smells using socio-technical metrics: A machine-learning approach," *Journal of Systems and Software*, p. 110847, 2020.

[29] G. Catolino, F. Palomba, D. A. Tamburri, and A. Serebrenik, "Understanding community smells variability: A statistical approach," in *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 2021, pp. 77–86.

[30] F. Shull, J. C. Carver, S. Vegas, and N. J. Juzgado, "The role of replications in empirical software engineering," *Empir. Softw. Eng.*, vol. 13, no. 2, pp. 211–218, 2008. [Online]. Available: https://doi.org/10.1007/s10664-008-9060-1

[31] "camila sarmento, tiago massoni, alexander serebrenik, gemma catolino, damian andrew tamburri and fabio palomba. 2021. the collected data and scripts (replications) - online appendix," https://zenodo.org/record/5669692.

[32] K.-J. Stol and B. Fitzgerald, "Guidelines for conducting software engineering research," in *Contemporary Empirical Methods in Software Engineering*. Springer, 2020, pp. 27–62.

[33] S. Baltes and P. Ralph, "Sampling in software engineering research: A critical review and guidelines," *arXiv preprint arXiv:2002.07764*, 2020.

[34] N. M. Razali, Y. B. Wah *et al.*, "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests," *Journal of statistical modeling and analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[35] T. W. MacFarland and J. M. Yates, "Kruskal–wallis h-test for oneway analysis of variance (anova) by ranks," in *Introduction to nonparametric statistics for the biological sciences using R*. Springer, 2016, pp. 177–211.

[36] J. Cohen, "Statistical power analyses for the behavioral sciences (rev. ed.) new york," 1977.

[37] T. Gnatzy, "Delphi studies and scenario planning: methodological advancements and cases," Ph.D. dissertation, EBS Business School, 2011.

[38] D. A. Tamburri, R. Kazman, and H. Fahimi, "Organisational structure patterns in agile teams: An industrial empirical study," *arXiv preprint arXiv:2004.07509*, 2020.

[39] K. Avrachenkov, G. Neglia, and A. Tuholukova, "Subsampling for chain-referral methods." in *ASMTA*, ser. Lecture Notes in Computer Science, S. Wittevrongel and T. Phung-Duc, Eds., vol. 9845. Springer, 2016, pp. 17–31. [Online]. Available: http://dblp.uni-trier.de/db/conf/asmta/asmta2016.html#AvrachenkovNT16