

# Diversidade de Gênero em Projetos Open Source no GitHub: um Estudo sobre a Relevância Temática dos Comentários das *Issues*

Estela Miranda Batista  
Universidade Federal de Viçosa  
Florestal, Brasil  
estela.batista@ufv.br

Gláucia Braga e Silva  
Universidade Federal de Viçosa  
Florestal, Brasil  
glauucia@ufv.br

Thais Regina de Moura  
Braga Silva  
Universidade Federal de Viçosa  
Florestal, Brasil  
thais.braga@ufv.br

## ABSTRACT

This work presents a comparative study of gender diversity on GitHub issue tracking based on metrics such as thematic relevance of comments, developer reputation, platform participation time and number of reported issues. Data were extracted and mined from 5 open source and 5 dedicated to women communities. The results show that, on average, the relevance of comments made by women is similar to those made by men. In connection with previous work from literature, the study also shows under-representation and low participation of women in software projects, with only 22% of comments posted by women and 16% of issues reported by them. Finally, the data showed a discrepancy between women and men platform participation time, being 2 to 9 years to the former, and 4 to 10 years or more to the latter, with a direct impact on the women's reputation.

## KEYWORDS

Female Representativeness and Participation, Reputation, Platform Participation Time

## 1 INTRODUÇÃO

Equipes de software podem ser formadas por pessoas de diversas culturas, idades, nacionalidades, gêneros, entre diversos outros fatores, de forma que aquelas que possuem uma maior diversidade social podem aproveitar de informações mais amplas, experiências e habilidades aprimoradas de resolução de problemas, tornando-se, portanto, mais eficazes [1]. Apesar do fator da diversidade de gênero ser benéfico para as equipes de desenvolvimento, existem trabalhos na literatura que apontam baixos índices da mesma no desenvolvimento de *software*, principalmente em questões de gênero, em que mulheres tendem a ter uma permanência menor que a de homens [2]. Porém, apesar de terem uma permanência menor nos projetos, estudos mostram que as mulheres funcionam como mediadoras dos problemas que surgem [3], sendo peças chave no processo de comunicação.

A baixa participação feminina em projetos de *software* é bem conhecida na literatura e tem sido avaliada, principalmente, em termos da contabilização das contribuições em código e *pull request* (solicitações de envio e publicação de código) aceitas [4, 5]. No entanto, outros aspectos podem ser analisados com o intuito de se investigar a participação feminina na composição de equipes mais diversas, mais inclusivas e, potencialmente mais produtivas [3]. Dentre os possíveis aspectos que podem ser analisados, destacam-se as comunicações em projetos de software. À medida que o desenvolvimento acontece, as interações das pessoas desenvolvedoras vão sendo registradas por meio de mensagens nas ferramentas de comunicação e

gerenciamento de tarefas, além das *issues* e comentários associados, em ambientes de *issue tracking*.

Este trabalho avalia as comunicações em *issue tracking*, por meio da aplicação da métrica da Relevância Temática [6] sobre os comentários postados. Essa métrica é capaz de representar o quanto um texto é relevante em relação ao tema de uma determinada discussão [7]. Neste trabalho, a relevância temática dos comentários será analisada em conjunto com outras métricas, como o número de *issues* criadas e número de comentários postados, em função de outros atributos que caracterizem o perfil da pessoa desenvolvedora como gênero, tempo de participação na plataforma e reputação. O estudo foi guiado pelas seguintes questões de pesquisa:

- **QP1.** Qual a diferença de participação de homens e mulheres em termos de *issues* e comentários associados?
- **QP2.** Existe diferença entre a relevância dos comentários postados por homens e mulheres?
- **QP3.** Existe relação entre a relevância dos comentários e o tempo de participação nas plataformas?
- **QP4.** Existe relação entre a relevância dos comentários e a reputação do autor?

Para responder às questões de pesquisa enumeradas, serão avaliados dados de comunicação extraídos de repositórios de projetos *open source* na plataforma do GitHub, em busca de análises comparativas entre a participação de homens e mulheres no contexto do *issue tracking*. O GitHub será usado por ser a maior comunidade de código aberto, possuindo mais de 61 milhões de repositórios de *software* criados por mais de 16 milhões de pessoas cadastrados na plataforma<sup>1</sup>, além de ser cada vez mais usado entre os pesquisadores como fonte de mineração de dados [8]. Para fins de comparação sobre a participação de mulheres e homens nesses ambientes, foram extraídos dados a partir de dois tipos de comunidades *open source*: comunidades abertas e comunidades dedicadas à mulheres.

Como resultados, serão apresentados indicativos importantes sobre o processo de comunicação dos times e as possíveis diferenças em termos de contribuições em função do gênero da pessoa desenvolvedora. Ressalta-se que o termo gênero é tratado a partir da forma de identificação do usuário na plataforma, ou seja, o nome atribuído, podendo assumir os valores masculino ou feminino. Não faz parte do escopo deste trabalho tratar outras identidades de gênero (transgêneros, cisgêneros e não-binários), uma vez que essa informação não está disponível diretamente na plataforma do GitHub.

O restante do artigo está organizado da seguinte forma: a Seção 2 apresenta informações sobre comunicações em *issue tracking*, e o conceitos de relevância temática dos comentários. Na Seção 3, são

<sup>1</sup><https://octoverse.github.com/>

apresentados os trabalhos relacionados. A Seção 4 traz os materiais e métodos utilizados no trabalho. Os resultados são discutidos na Seção 5. As ameaças à validade do trabalho foram descritas na Seção 6, e na Seção 7 são apresentadas as considerações finais.

## 2 COMUNICAÇÕES EM ISSUE TRACKING

Ao longo do desenvolvimento de *software* surgem diversas questões sobre melhorias e dúvidas apontadas pelos membros dos projetos. Essas questões ficam comumente concentradas em tópicos de discussão, representados por *issues* no sistema de *issue tracking* do GitHub. Segundo Bertran et al. [9], os repositórios de *issue tracking* são repositórios de conhecimento, que concentram grande parte da comunicação e da colaboração dos projetos gerando conhecimento útil e relevante.

Análises automatizadas sobre esses dados podem revelar indicadores de qualidade de processo importantes no que diz respeito à participação dos membros nas discussões dos projetos, como, por exemplo, pessoas desenvolvedoras-chaves, relevância dos comentários, conflitos nas relações, representatividade e participação de novatos e minorias. Para auxiliar nesse processo de análise existem algumas métricas usadas para tal apontamento, como a reputação dos usuários, a contabilização dos anos de participação do usuário na plataforma, e a qualidade dos comentários postados.

### 2.1 Issue Tracking no GitHub

Nos ambientes de *issue tracking*, as comunicações ocorrem à medida que *issues*, representando *bugs* (defeitos, erros ou falhas), sugestões de melhoria e novos requisitos, são reportadas pelos membros dos projetos e/ou por colaboradores externos. Após reportadas, as *issues* podem ser atendidas e/ou discutidas por qualquer membro da equipe ou por colaboradores externos. Ao final da discussão em busca da solução, podem ser gerados novos *commits* (formas de enviar as alterações mais recentes do código-fonte ao repositório), e consequentemente *pull requests* avaliadas por gerentes de projeto, ou por algum responsável pelo repositório.

As *issues* são compostas por alguns campos obrigatórios, e outros, que dependendo do projeto não necessitam ser preenchidos. O campo obrigatório é o título. Já os campos opcionais incluem a descrição, *labels* (rótulos), como *bug* ou *question*, e a quem aquela *issue* é atribuída. Por fim, tem-se os campos gerados automaticamente pela plataforma, como o autor, data de criação, e o estado (*open* ou *closed*). Os campos relacionados a uma *issue* podem ser observados na Figura 1<sup>2</sup>.

Os campos *título* e *descrição* contêm a principal fonte de dados textuais, uma vez que representam o conteúdo da discussão realizada na *issue* através dos comentários posteriormente postados, e, por conta disso, serão explorados no contexto deste trabalho.

### 2.2 Relevância Temática em Issue Tracking

A métrica da relevância temática foi proposta por Azevedo [10] para se analisar a relevância de comentários postados em fóruns de discussão educacionais, e adaptada para o ambiente de *issue tracking* por Neto e Braga e Silva [6]. O objetivo de utilizar a relevância temática no contexto do *issue tracking* é identificar se o texto do

comentário está relacionado ao contexto do tema proposto na *issue*, com o intuito de se avaliar o impacto do comentário na resolução da *issue*. Com essa métrica, também é possível investigar se existem correlações com outros atributos ligados à pessoa desenvolvedora, como gênero, reputação e o tempo de projeto.

O cálculo desta métrica envolve contabilizar o número de conceitos relevantes usados no texto do comentário que correspondem a conceitos do texto da problematização, composto pelo título e pela descrição da *issue*. Além de contabilizar os conceitos relevantes, o cálculo também considera a frequência dos mesmos no texto e os relacionamentos com outros termos. Assim, quanto mais conceitos relacionados, maior será a relevância do comentário. A Equação 1, adaptada por Neto e Braga e Silva [6], apresenta o cálculo da relevância temática.

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

Onde *RT* é a Relevância Temática de um comentário; *S<sub>CI</sub>*, a similaridade entre o comentário e a *issue*, sendo representada pelo seu título e descrição; *S<sub>CD</sub>*, a similaridade entre o comentário e a discussão, que leva em consideração o título da *issue*, sua descrição e o comentário anterior ao analisado; e *S<sub>CC</sub>*, a similaridade entre o comentário e o comentário anterior, caso exista. O valor de *RT* é dado pelo maior valor de similaridade encontrada, como apresentado na Equação 1.

Para realizar o cálculo das similaridades apresentadas, duas técnicas podem ser utilizadas: grafos e cossenos. Como apresentado por Machado et al. [7], na técnica de grafos, os textos de um comentário são representados em um grafo, em que os termos mais relevantes constituem os vértices e as arestas conectam os termos que aparecem em conjunto no texto. Esse grafo é então comparado com um outro grafo, construído de forma similar, mas com base no texto da *issue*. Para comparação entre os grafos, a similaridade é calculada a partir da correspondência entre os termos. Já a técnica de cossenos propõe que cada texto seja vetorizado, para que possa ser realizada uma análise angular entre os conteúdos textuais [11]. Assim, ao se comparar o vetor de um comentário com o vetor da *issue*, por exemplo, realiza-se o cálculo do cosseno do ângulo entre os dois vetores que representam as interações, gerando assim a similaridade.

## 3 TRABALHOS RELACIONADOS

A baixa participação feminina em projetos de desenvolvimento de *software* tem sido reportada na literatura. Zacchiroli [12] analisou 1.6 milhões de *commits* e apresentou que cerca de 92% do código produzido nas plataformas de versionamento de código foi feito por homens. Izquierdo et al. [13] analisaram mais de 7000 perfis de usuários na plataforma do GitHub, em que apenas 7% dos membros do comitê de projeto e 8% dos líderes foram identificados como mulheres. Buscando apresentar novos indicadores sobre a baixa participação feminina no desenvolvimento de *software*, este trabalho visa analisar a quantidade e a qualidade do conteúdo das comunicações em *issue tracking*.

No que compete à análise dos dados de comunicações em *issue tracking*, trabalhos na literatura têm usado técnicas de mineração de textos sobre o conteúdo textual das *issues*. Ortu et al. [14] analisou os dados de comunicação em projetos de *software* com o objetivo de

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



Figura 1: Exemplo de *Issue* no GitHub

avaliar o impacto de determinados fatores no tempo de resolução de uma *issue*. No trabalho de Neto e Braga e Silva [6], foi extraído conhecimento sobre pessoas desenvolvedoras do projeto, indicando pessoas desenvolvedoras chaves no projeto, usando de métricas como a relevância temática dos comentários, o número de *issues* reportadas e o número de comentários postados nas *issues*. No entanto, apesar de trazerem investigações pertinentes às comunicações do desenvolvimento de *software*, os trabalhos apresentados não levam em consideração questões de gênero.

Noei and Lyons [15], por sua vez, avaliaram cerca de 700 mil comentários de *review* de aplicativos na plataforma da Google Play Store, em que foi possível notar que as mulheres realizaram menos comentários que os homens. Além disso, os autores também observaram que a maior parte dos comentários postados por mulheres são positivos, pois demonstram elogios aos aplicativos, ao passo que os dos homens trazem críticas e sugestões de melhorias. Isso faz com que os ajustes e novas versões gerados em função dos comentários reflitam, em sua maioria, a opinião masculina.

Por fim, considerando a participação feminina no universo *open source*, Singh [16] analisou 355 sites de projetos, e os resultados apontam que menos de 5% dessas comunidades possuíam espaço dedicados às mulheres. Neste trabalho, serão utilizados ambientes dedicados para mulheres para fins de comparações com ambientes com baixa diversidade, indicados pelo índice Blau [17], de forma a verificar se as mulheres são mais atuantes em termos de comunicações em ambientes dedicados a elas.

## 4 MATERIAIS E MÉTODOS

Para analisar os dados de comunicação em *issue tracking* em função do gênero da pessoa desenvolvedora, foram executadas duas etapas: definição das métricas e extração dos dados.

### 4.1 Definição e Automatização das Métricas

Para a condução das análises, foram utilizadas seis métricas gênero-dependentes, chamadas aqui de Métricas de Comunicação (MC).

- **MC01.** Quantidade de *Issues* Reportadas;
- **MC02.** Quantidade de Comentários Postados;

- **MC03.** Tempo de Participação na Plataforma;
- **MC04.** Reputação da Pessoa Desenvolvedora;
- **MC05.** Índice de Diversidade de Gênero das Equipes [17];
- **MC06.** Relevância Temática dos Comentários.

Para obter as medidas para as métricas MC01 a MC04, foram utilizadas APIs e ferramentas da plataforma do GitHub. Já para as métricas MC05 e MC06, foi desenvolvida uma aplicação em Python, conforme descrito no decorrer desta seção.

As métricas MC01 e MC02 foram calculadas de acordo com o gênero da pessoa desenvolvedora. Após identificado o gênero de cada pessoa desenvolvedora através da ferramenta NamSor<sup>3</sup>, os dados foram armazenados em um arquivo CSV, e utilizando de filtros no Jupyter Notebook<sup>4</sup> foram gerados tais cálculos. Já a métrica MC03 foi calculada de acordo com a data fornecida pela API GitHub, sendo subtraído o valor do ano de entrada da pessoa desenvolvedora pelo ano corrente (2022).

Para a métrica MC04, visto que a plataforma do GitHub não realiza a contabilização automática da reputação das pessoas desenvolvedoras, foi utilizada a ferramenta GitScore<sup>5</sup>. A ferramenta possui três campos: GitScore, Reputação e Contribuição, sendo neste trabalho considerado o campo reputação. Esse campo leva em consideração as conexões da pessoa desenvolvedora, sendo usada como um atributo social, para complementar outras métricas de contribuições das pessoas desenvolvedoras aplicadas, como quantidade de comentários e relevância temática associada.

No que diz respeito à diversidade de gênero das equipes (MC05), foi usado o Índice Blau como métrica de diversidade [17]. A métrica de diversidade foi usada com o propósito de realizar comparações sobre os resultados das métricas aplicadas aos dados de comunicação para cada um dos segmentos de comunidades avaliadas. A Equação 2 apresenta a fórmula para o índice Blau, que calcula, de um total de  $N$ , a porcentagem  $P$  de indivíduos em cada categoria  $i$ . Neste trabalho, consideramos  $N = 2$ , visto que as categorias são masculino e feminino. O índice varia entre 0 e 0,5, sendo 0,5 o equilíbrio no número de indivíduos nas categorias.

<sup>3</sup><https://namsor.app/>

<sup>4</sup><https://jupyter.org/>

<sup>5</sup><http://www.git-score.com>

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

Ao final da extração dos dados de uma determinada comunidade, o nome e o gênero das pessoas desenvolvedoras participantes foi salvo em um arquivo CSV. A partir desse arquivo, usando dos filtros do Jupyter Notebook no arquivo CSV gerado, foi contabilizado o número de pessoas desenvolvedoras nos gêneros feminino e masculino. Tais valores foram enviados a uma função simples na linguagem Python, que automatizou a métrica MC05, retornando assim o nível de diversidade daquela comunidade.

Visto que na plataforma de *issue tracking* do GitHub não é obrigatória a identificação do gênero das pessoas desenvolvedoras, e as métricas anteriormente apresentadas fazem comparações em relação ao gênero, fez-se necessário "adivinhar" tal informação. Zolduoarrati [18] realizou um estudo para qualificar as melhores ferramentas adivinhadoras de gênero. Dentre as ferramentas apresentadas, após uma avaliação para o contexto deste trabalho, foi escolhida a ferramenta NamSor, devido à sua simplicidade de uso e a qualidade dos resultados. Para a adivinhação, basta fornecer o nome e o sobrenome de uma determinada pessoa desenvolvedora, que a ferramenta retornará uma porcentagem de chance do usuário possuir o gênero feminino ou masculino, definindo-se o gênero pela maior porcentagem. Além disso, a ferramenta possui suporte à diversos idiomas, como português, inglês, japonês e chinês, bastando que seja inserido o nome e sobrenome de uma pessoa.

## 4.2 Relevância Temática dos Comentários

Para a avaliação da qualidade dos comentários (MC06), foi desenvolvida uma aplicação automatizada da métrica de relevância temática (Seção 2.2). Para fins de simplificação e melhoria no desempenho, o cálculo automatizado da métrica foi adaptado no que diz respeito ao cálculo da similaridade entre os termos de um comentário e seu texto referência usando da técnica de cossenos. Na proposta de Neto e Braga e Silva [6], o cálculo da similaridade é realizado através de grafos, com uma dependência externa a uma ferramenta de geração dos mesmos. Essa dependência traz alguns problemas de desempenho em virtude do alto número de requisições. A alteração proposta neste trabalho propiciou uma melhora significativa no desempenho da aplicação, sem provocar alterações em termos da qualidade dos resultados gerados, conforme detalhado na Seção 4.2.1. A técnica de cossenos foi escolhida também devido à simplicidade de aplicação e também por apresentar uma alta precisão em relação à classificação humana, como observado por Medeiros [19].

Ao usar a fórmula da Equação 1, apresentada na Seção 2.2, observou-se que a similaridade do comentário em relação ao seu comentário anterior ( $S_{CC}$ ) assumia, em diversos momentos, o valor zero, mostrando-se irrelevante para o cálculo da relevância temática. Assim, após alguns testes, foi possível notar que a similaridade do comentário em relação à discussão ( $S_{CD}$ ) e a similaridade do comentário em relação à *issue* ( $S_{CI}$ ) eram suficientes para compor a equação de cálculo da relevância temática. Optou-se então pela média aritmética de  $S_{CI}$  e  $S_{CD}$ , como apresentado na Equação 3.

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

**4.2.1 Validação da Adaptação no Cálculo da Métrica.** Para validar a alteração e garantir a confiabilidade dos resultados, foi realizado um procedimento de validação, comparando os resultados da versão original [6] e da versão atualizada com uma avaliação manual de especialistas. O grupo de especialistas contou com um total de 12 voluntários, sendo 3 engenheiros de *software*; 2 especialistas de domínio, ou seja, que participam da *issue* em questão; e 7 pessoas desenvolvedoras, com níveis de conhecimento variados.

Para a avaliação de especialistas, foi elaborado um questionário a ser respondido para cada *issue*, onde cada especialista atribuiu uma relevância de 0 a 4 para cada comentário daquela *issue*. Embora a relevância seja calculada no intervalo de 0 a 1, a escala de 0 a 4 foi usada para simplificar a atribuição pelos especialistas e evitar eventuais erros com as casas decimais. No entanto, posteriormente, os valores foram convertidos para a mesma escala. Ao todo foram analisados 59 comentários, distribuídos em 12 *issues*. Cada *issue* foi analisada por dois especialistas, de forma que cada especialista atribuiu valores de relevância para cada um dos comentários registrados para aquela *issue*. Juntamente à avaliação dos especialistas, foi calculada uma média com as relevâncias atribuídas pela equipe de desenvolvimento da métrica. A Tabela 1 apresenta um exemplo do formato de avaliação de um determinado comentário, em que a relevância final corresponde à média aritmética entre os valores de relevância definidos pelos especialistas e a média de relevância atribuída pela equipe de desenvolvimento.

**Table 1: Exemplo de Avaliação Manual de Relevância Temática dos Comentários de uma Issue**

Comentário	#Média Equipe	#Especialista 1	#Especialista 2	#Relevância Atribuída
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

Para verificar a precisão do cálculo automatizado após as adaptações da abordagem, em comparação com a classificação dos especialistas e com a versão original [6], foram usadas as métricas MAE (*Mean Absolute Error* ou Erro Absoluto Médio) e MSE (*Mean Squared Error* ou Erro Quadrático Médio). O cálculo de MAE (Equação 4), representa a média da diferença absoluta entre os valores reais e os previstos no conjunto de dados. A equação considera o valor da relevância final dos especialistas ( $y$ ) e os valores calculados automaticamente ( $\hat{y}$ ), seja pela abordagem original ou pela abordagem adaptada, ou seja, o cálculo foi aplicado duas vezes, uma para cada abordagem considerada.

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

O cálculo de MSE representa a média quadrada entre os valores reais e os previstos no conjunto de dados, como apresentado na Equação 5.

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

Ao final, observou-se uma melhora nos valores de MSE e MAE, para a versão adaptada, em comparação com a abordagem original [6]. A porcentagem de erro da abordagem original ficou em torno de 21% no MAE e 7% no MSE, enquanto na abordagem adaptada, houve redução do valor de MAE, para 8% e de MSE, para 1%.

### 4.3 Seleção e Extração dos Dados

Na seleção das bases de dados, foram considerados projetos de comunidades abertas, e comunidades dedicadas às mulheres, de forma a se comparar a participação feminina nesses dois contextos. As comunidades dedicadas foram selecionadas através de uma pesquisa preliminar, em que após a escolha de comunidades abertas, buscou-se comunidades no mesmo segmento que trabalhassem a inclusão de mulheres.

Os dados foram extraídos em janeiro de 2022, por meio de uma aplicação na linguagem Python, utilizando a API Rest do GitHub<sup>6</sup>.

Para a etapa de extração dos dados, foram utilizados filtros baseados no trabalho de Neto et al. [20], divididos em duas categorias: Filtros de Projetos (FPs) e Filtros de *Issues* (FIs). Para selecionar os projetos de interesse, foram definidos cinco filtros *FPs*:

- **FP1.** Possuir, no mínimo, 5 membros;
- **FP2.** Possuir, no mínimo, 5 *commits*;
- **FP3.** Possuir, no mínimo, 5 *issues* abertas;
- **FP4.** Possuir, no mínimo, 5 *issues* fechadas;
- **FP5.** Ter sido criado há pelo menos 6 meses.

A seleção das *issues* abrangeu a aplicação dos seguinte filtros *FIs*:

- **FI1.** Possuir no mínimo, 5 comentários.;
- **FI2.** Os comentários não podem conter apenas trechos de códigos, sendo assim necessário textos;
- **FI3.** Possuir um tempo de abertura mínimo de uma semana;
- **FI4.** *Issues* não podem ter sido reabertas.

Ressalta-se que o filtro FP4 foi definido com base no fator de que, as *issues* avaliadas no trabalho estavam com seu *status closed*, de forma que pudéssemos garantir que ao longo da extração dos dados, e da pesquisa não fossem inseridos mais comentários, unido a isso temos a definição do filtro FI4, para trazer mais uma garantia. Já o filtro FP5 foi escolhido de forma a garantir que o projeto estivesse plenamente em funcionamento, e executando as tarefas pertinentes ao sistema de *issue tracking*, unido aos fatores apresentados nos filtros FP2, FP3 e FP4. O filtro FI2 foi inserido com base no fator de que a avaliação dos comentários é feita a partir de seus textos, ou seja, os trechos de código não são inseridos na avaliação da relevância temática, fazendo com que comentários com apenas trechos de código se tornem irrelevantes. Além disso, é importante ressaltar que o filtro FI2 leva em consideração todos os comentários de uma determinada *issue*, de forma que, caso tenhamos um comentário com apenas trechos de códigos, aquela *issue* será desconsiderada.

Ao final, esses dados foram armazenados em um arquivo CSV, que posteriormente foram analisados através do Jupyter Notebook.

<sup>6</sup><https://docs.github.com/pt/rest>

## 5 RESULTADOS E DISCUSSÕES

Após a extração dos dados com aplicação dos filtros, foram obtidos 9151 comentários, de 1275 *issues*, presentes em 28 repositórios de 10 comunidades, sendo 5 comunidades abertas e 5 dedicadas às mulheres. Na Tabela 2, tem-se um sumário dos dados obtidos para cada comunidade, incluindo o número de repositórios, *issues* e comentários. Ressalta-se que as cinco primeiras comunidades são dedicadas às mulheres.

Table 2: Visão Geral dos Dados Extraídos por Comunidades

Comunidade	#Repositórios	#Issues	#Comentários
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

Para cada uma das comunidades, foi calculado o quantitativo de mulheres e homens participantes e também o índice Blau, com o intuito de se conhecer o fator de diversidade em cada uma delas. Os resultados são apresentados na Tabela 3, em que é possível observar as discrepâncias em termos de representatividade feminina nas comunidades abertas e, por consequência, o baixo índice de diversidade de gênero.

Table 3: índice Blau e Quantitativos por Gênero

Comunidade	#Mulheres	#Homens	#Índice Blau
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

Conforme esperado, os dados coletados evidenciam baixos índices de diversidade em comunidades abertas e índices altos em comunidades dedicadas, já que o maior índice chegou a 0.5, indicando o equilíbrio entre as categorias, na RailsGirls. Importante também destacar que na comunidade PyLadies, apesar do índice Blau ser igual a 0.0 indicar a falta de diversidade, isso ocorre porque a comunidade é fechada à participação masculina, com o intuito de se criar um ambiente seguro para que as mulheres possam desenvolver as habilidades, tanto técnicas, quanto teóricas. Ao se observar a Tabela

2 e a Tabela 3, comunidades como o Ruby on Rails tem um alto número de comentários postados, porém, seu nível de diversidade é baixo. Importante também ressaltar, que após uma análise na plataforma do GitHub, de forma a analisar os nomes e sobrenomes dos membros das comunidades, foi possível notar que muitas mulheres que estavam presentes na comunidade Ruby on Rails também se faziam presentes na comunidade de mesmo tema, RailsGirls. Isso pode indicar que a participação na comunidade aberta tenha sido influenciada pelo empoderamento e pela segurança adquiridos ao participarem da comunidade dedicada.

Os demais resultados obtidos com a aplicação das métricas MC01 a MC06, voltadas às questões de pesquisa endereçadas neste estudo, encontram-se descritos nas próximas subseções.

### 5.1 QP1. Qual a diferença de participação de homens e mulheres em termos de *issues* e comentários associados?

Olhando primeiramente para as *issues*, foi possível identificar o gênero de todas as pessoas envolvidoras através da ferramenta de adivinhação de gênero, sem qualquer intervenção manual, sendo extraídas 1275, em que 1071 (84%) foram reportadas por homens, e 204 (16%) por mulheres, como apresentado na Tabela 4.

Table 4: Quantitativos por Gênero de Issues e Comentários

Comunidade	Issues		Comentários	
	#Mulheres	#Homens	#Mulheres	#Homens
Dedicada	102	116	741	1166
Aberta	102	955	1395	5872

Considerando o recorte das comunidades dedicadas, tem-se 218 *issues* reportadas, das quais 102 (48%) foram reportadas por mulheres. Já em comunidades abertas, foram reportadas 1057 *issues*, sendo que apenas 102 (9%) foram reportadas por mulheres. Tais valores mostram que com o aumento de mulheres nas comunidades, a porcentagem de participação das mesmas aumenta. Ao contrário dos homens, que possuem altos valores em ambas as comunidades.

No que compete aos comentários postados para as *issues*, foi possível obter o gênero dos autores automaticamente para a maioria dos registros. Em alguns casos, foi necessária a intervenção manual, em que apenas uma pessoa desenvolvedora permaneceu com o gênero indefinido. A intervenção manual foi realizada de forma a filtrar os usuários com gênero não identificado através do Jupyter Notebook. Nesses casos, o *nickname* dos usuários foi usado como campo de pesquisa na plataforma do GitHub, em busca de informações complementares que pudessem ajudar na identificação, tais como, imagem de perfil do usuário, descrição, e quando possível blogs pessoais linkados. Além disso, tem-se que alguns comentários foram postados por *bots*, não sendo, assim, considerados na Tabela 4. A partir disso, dos 9151 comentários avaliados, 6897 foram postados por homens, 2059 (22%) por mulheres e 194 por *bots*.

No contexto das comunidades dedicadas, dos 1923 comentários, 741 (39%) foram postados por mulheres. E no caso das comunidades abertas, foram postados 7447 comentários, em que apenas 1395

(19%) foram postados por mulheres, o que também evidencia a baixa participação em comparação com os homens neste segmento.

### 5.2 QP2. Existe diferença entre a relevância dos comentários postados por homens e mulheres?

Após o cálculo da relevância temática para cada comentário analisado, os resultados são similares na média, sendo 0.03680, para os homens e 0.03394, para as mulheres. Conforme mostrado no gráfico das Figuras 2 e 3, observa-se que o intervalo de dados altera entre os gêneros, segundo o contexto do segmento em análise. Nas comunidades dedicadas (Figura 2), diferente das abertas (Figura 3), as mulheres alcançam valores de relevância maiores que os homens.

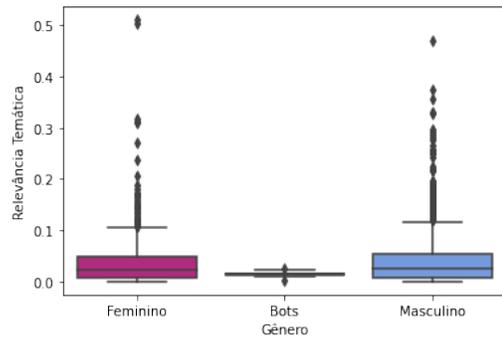


Figura 2: Relevância Temática nas Comunidades Dedicadas

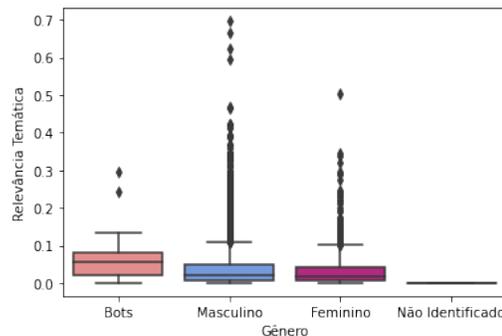


Figura 3: Relevância Temática nas Comunidades Abertas

Os resultados apresentados demonstram que apesar das mulheres possuírem uma quantidade menor de comentários postados, como apresentado na Seção 5.1, independente da comunidade, aberta ou dedicada, seus comentários mostraram-se igualmente relevantes para a discussão em torno das *issues*.

### 5.3 QP3. Existe relação entre a relevância dos comentários e o tempo de participação na plataforma?

Para responder a essa primeira questão de pesquisa, os dados de relevância temática e o tempo de participação na plataforma foram

usados para construir o gráfico da Figura 4, em que é possível observar valores de relevância temática muito similares entre si.

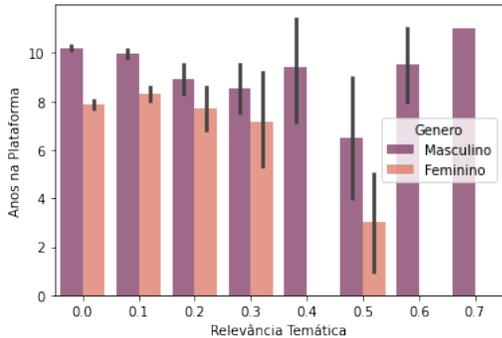


Figura 4: Relevância Temática e Tempo de Plataforma

Para avaliar melhor essa observação, aplicou-se a correlação de Pearson [21]. O coeficiente de Pearson varia de -1 a 1, indicando uma tendência crescente quando positivo e próximo de 1. Para os dados analisados, o valor da correlação entre a relevância temática dos comentários e o tempo de participação na plataforma foi de -0.052, indicando assim a falta de correlação entre essas variáveis.

É importante ressaltar que as linhas destacadas em preto na Figura 4 estão indicando o intervalo em que 50% ou mais dos dados estão presentes, trazendo outro indicativo da falta de correlação, visto que valores baixos de relevância temática possuem comentários feitos por pessoas desenvolvedoras com 10 anos ou mais.

Por fim, também é possível observar que os intervalos que representam o tempo de participação das mulheres é menor do que o dos homens. O intervalo de participação das mulheres está variando de 2 até cerca de 9 anos, enquanto o dos homens está variando de 4 anos até mais de 10 anos. Tais valores são refletidos tanto em comunidades abertas, quanto em comunidades dedicadas, ou seja, independente do contexto, as mulheres possuem uma participação mais recente e menor do que em relação aos homens.

#### 5.4 QP4. Existe relação entre a relevância dos comentários e a reputação do autor?

Os dados de relevância temática e reputação dos autores dos comentários estão projetados no gráfico da Figura 5. Observa-se que todos os valores de relevância temática possuem variações da reputação da pessoa desenvolvedora que postou o comentário, indicando que não parece haver relação entre essas duas variáveis. Além disso, o intervalo de reputação presente na relevância de 0.0 está em cerca de  $10^3$ , e a reputação presente na relevância de 0.5 chega até aproximadamente  $10^2$ , como indicado pela linha preta no gráfico.

Novamente, para confirmar essa observação, foi feito o cálculo do coeficiente de Pearson entre a relevância temática dos comentários e a reputação dos autores de comentário, resultando em -0.045, novamente uma baixa correlação entre essas duas variáveis.

Na Figura 5, também é possível notar que as mulheres possuem um intervalo de reputação menor do que o dos homens, isso porque o intervalo destacado pela linha preta de colunas que representam os dados de mulheres chega até quase o valor de  $10^3$  em apenas um

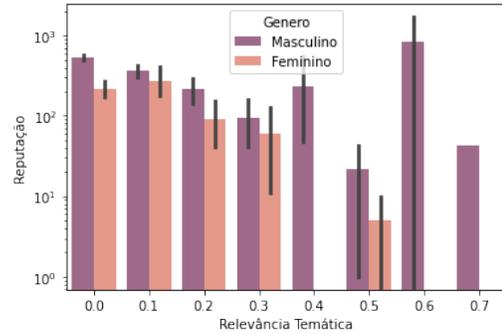


Figura 5: Relevância Temática e Reputação para Mulheres

caso, enquanto o intervalo de reputação dos homens em mais de um caso quase chega ao valor de  $10^3$ , e em outros chegam até mesmo a passar esse valor. Tal questão pode ser justificada ao se analisar em conjunto as Figuras 4 e 5, em que o tempo de plataforma tem um impacto na reputação das pessoas desenvolvedoras. Para confirmar essa análise, foi calculado o coeficiente de Pearson, obtendo-se um valor 0.249, que indica haver uma correlação positiva entre essas variáveis. Ou seja, como as mulheres possuem um menor tempo de participação na plataforma do GitHub, chegando, no máximo, a 9 anos, poderão ter uma menor reputação, já que consequentemente farão menos comentários, *commits*, e demais atividades, elementos usados no cálculo da reputação. É importante ressaltar que, tanto em comunidades dedicadas quanto em comunidades abertas, esse cenário se repete, porém, com a ressalva de que as mulheres acabam possuindo uma reputação similar a dos homens em comunidades dedicadas, em que ambos não chegam ao valor de  $10^3$ , conforme apresentado na Figura 6.

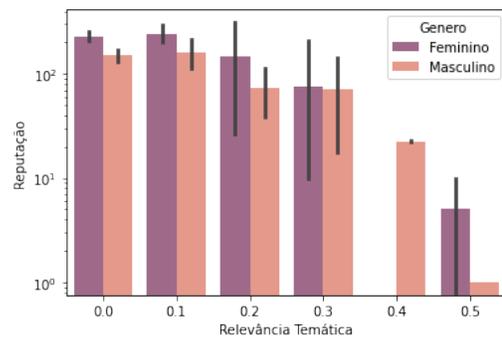


Figura 6: Relevância Temática e Reputação para Mulheres em Comunidades Dedicadas

## 6 AMEAÇAS À VALIDADE

Este trabalho considerou apenas gêneros feminino e masculino das pessoas desenvolvedoras. Essa decisão foi tomada para fins de simplificação deste primeiro estudo e também em função das limitações da ferramenta NamSor, que usa o nome para a adivinhação.

Em relação ao uso da métrica da relevância temática, ressalta-se que ainda existem poucos estudos referentes à sua aplicação no

contexto de *issue tracking*. Além disso, a natureza dos comentários das *issues*, que podem conter imagens, trechos de código, links externos, mensagens de agradecimento, entre outros elementos, pode resultar em valores relativamente baixos de relevância, seja porque este conteúdo é ignorado pela métrica ou por não gerar similaridade com a temática da *issue*.

Por fim, em relação ao uso da ferramenta GitScore, para o cálculo da métrica MC04, não foi realizado um procedimento de validação para verificar a precisão dos resultados obtidos.

## 7 CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma análise de dados de *issue tracking* com foco em diversidade de gênero. Observou-se que, na média, as mulheres postam comentários com relevância similar aos dos homens. O estudo reforçou que os índices discrepantes de participação feminina nas comunidades *open source* também se confirmam em ambientes de *issue tracking*, com apenas 22% dos comentários postados por mulheres e 16% das *issues* com autoria feminina. Olhando para os recortes das comunidades, verificou-se que as mulheres têm sido mais atuantes em comunidades dedicadas à elas, chegando a 48% das *issues* reportadas. Porém, nas comunidades abertas, observou-se baixos índices de participação e representatividade, com apenas 9% das *issues* reportadas. Outro ponto importante diz respeito à discrepância no tempo de participação na plataforma, sendo 2 a 9 anos, para as mulheres, e 4 a 10 anos ou mais, para homens. Os dados disponibilizados podem ser úteis para gestores de projetos para fins de identificação e tratamento dos débitos sociais revelados no que compete à desigualdade de gênero nas organizações.

Como trabalhos futuros, levando-se em consideração questões técnicas sobre a análise de dados, pode-se realizar o cruzamento entre os dados da relevância temática com outros indicadores de participação no desenvolvimento de *software*, como o número de *commits* realizados e/ou *pull requests* aceitas. Pode-se avaliar se a participação feminina no âmbito das comunicações nos projetos causa algum impacto em termos do tempo de resolução ou fechamento de *issues*. Por fim, com o intuito de complementar a análise quantitativa dos dados, podem ser realizadas entrevistas com mulheres que participam das comunidades, para se conhecer as percepções das mesmas quanto aos motivos que levam aos baixos índices obtidos.

## DISPONIBILIDADE DE ARTEFATOS

Os artefatos utilizados no desenvolvimento da pesquisa encontram-se disponíveis em um repositório no GitHub<sup>7</sup>.

## REFERENCES

- [1] Bogdan Vasilescu, Alexander Serebrenik, and Vladimir Filkov. 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, 2015. IEEE. doi: 10.1109/MSR.2015.77.
- [2] Balazs Vedres and Orsolya Vasarhelyi. Gendered behavior as a disadvantage in open source software development. *EPJ Data Science*, 8, 12 2019. ISSN 21931127. doi: 10.1140/epjds/s13688-019-0202-z.
- [3] Gemma Catolino, Fabio Palomba, Damian A. Tamburri, Alexander Serebrenik, and Filomena 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*, pages 11–20, Montreal, QC, Canada, 2019. IEEE. doi: 10.1109/ICSE-SEIS.2019.00010.
- [4] Gema Rodriguez, Reza Nadri, and Meiyappan Nagappan. Perceived diversity in software engineering: a systematic literature review. *Empirical Software Engineering*, 26, 09 2021. doi: 10.1007/s10664-021-09992-2.
- [5] E.D. Canedo, Heloíse Tives, Madianita Bogo, Fabiano Fagundes, and José Siqueira de Cerqueira. Barriers faced by women in software development projects. *Information*, 10:309, 10 2019. doi: 10.3390/info10100309.
- [6] Luiz Eugênio Coelho Neto and Gláucia Braga e Silva. 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, 2018. Association for Computing Machinery. ISBN 9781450365598. doi: 10.1145/3229345.3229398. URL <https://doi.org/10.1145/3229345.3229398>.
- [7] Crystiano José Richard Machado, Alexandre Magno Andrade Maciel, Rodrigo Lins Rodrigues, and Ronaldo Menezes. An approach for thematic relevance analysis applied to textual contributions in discussion forums. *International Journal of Distance Education Technologies*, 17:37–51, 2019. ISSN 15393119. doi: 10.4018/IJDET.2019070103.
- [8] Samaneh Saadat, Olivia B. Newton, Gita Sukthakar, and Stephen M. Fiore. 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*, pages 169–176, Melbourne, Australia, 2020. IEEE. doi: 10.1109/WIAT50758.2020.00027.
- [9] Dane Bertram, Amy Voids, Saul Greenberg, and Robert Walker. 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, 2010. ACM. ISBN 978-1-60558-795-0. doi: 10.1145/1718918.1718972. URL <http://doi.acm.org.ez35.periodicos.capes.gov.br/10.1145/1718918.1718972>.
- [10] Breno Fabricio Terra Azevedo. 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, 2011.
- [11] Maicom Sergio Brandão, Moacir Godinho-Filho, Walther Azzolini Junior, Bruna Christina Battissacco, and Josadak Astorino Marçola. 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, mar. 2022. doi: 10.14488/1676-1901.v21i4.4483. URL <https://www.producaoonline.org.br/rpo/article/view/4483>.
- [12] Stefano Zacchiroli. Gender differences in public code contributions: A 50-year perspective. *IEEE Software*, 38(2):45–50, 2021. doi: 10.1109/MS.2020.3038765.
- [13] Daniel Izquierdo, Nicole Huesman, Alexander Serebrenik, and Gregorio Robles. Openstack gender diversity report. *IEEE Software*, 36:28–33, 1 2019. ISSN 19374194. doi: 10.1109/MS.2018.2874322.
- [14] Marco Ortu, Tracy Hall, Michele Marchesi, Roberto Tonelli, David Bowes, and Giuseppe Destefanis. 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, 2018. Association for Computing Machinery. ISBN 9781450365932. doi: 10.1145/3273934.3273943. URL <https://doi.org/10.1145/3273934.3273943>.
- [15] Ehsan Noei and Kelly Lyons. A study of gender in user reviews on the google play store. *Empirical Softw. Engg.*, 27(2), mar 2022. ISSN 1382-3256. doi: 10.1007/s10664-021-10080-8. URL <https://doi.org/10.1007/s10664-021-10080-8>.
- [16] Vandana Singh. Women-only spaces of open source. In *1999 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering*, pages 17–20, Montreal, QC, Canada, 2019. IEEE. doi: 10.1109/GE.2019.00010.
- [17] Torsten Biemann and Eric Kearney. Size does matter: How varying group sizes in a sample affect the most common measures of group diversity. *Organizational Research Methods*, 13:582–599, 7 2010. ISSN 10944281. doi: 10.1177/1094428109338875.
- [18] Elijah Zolduoarrati and Sherlock A. Licorish. On the value of encouraging gender tolerance and inclusiveness in software engineering communities. *Information and Software Technology*, 139:106667, 11 2021. ISSN 09505849. doi: 10.1016/j.infsof.2021.106667.
- [19] Danielle C. Medeiros, José Eustáquio Rangel de Queiroz, and Joseana M. F. R. Araújo. Análise de funções de similaridade para verificação do conteúdo de mensagens em fóruns de discussão. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação)*, number 1, page 144, Dourados, Mato Grosso do Sul, Brazil, 2014. SBC. doi: 10.5753/cbie.sbie.2014.144. URL <https://www.br-ie.org/pub/index.php/sbie/article/view/2940>.
- [20] Luiz Neto, Gláucia Silva, and Giovanni Comarela. 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, 2021. Association for Computing Machinery. ISBN 9781450390613. URL <https://doi.org/10.1145/3474624.3474647>.
- [21] Ranulfo Paranhos, Dalson Britto Figueiredo Filho, Enivaldo Carvalho da Rocha, José Alexandre da Silva Júnior, Jorge Alexandre Barbosa Neves, and Manoel Leonardo Wanderley Duarte Santos. Desvendando os mistérios do coeficiente de correlação de pearson: o retorno. *Leviathan (São Paulo)*, (8):66–95, ago. 2014. doi: 10.11606/issn.2237-4485.lev.2014.132346. URL <https://www.revistas.usp.br/leviathan/article/view/132346>.

<sup>7</sup><https://github.com/stardotwav/AnaliseGeneroGitHub>