

Perfil dos Contribuidores de Projetos de Código Aberto: Uma Análise de 1000 Projetos das Linguagens Mais Populares do GitHub

Arthur F. Costa¹, Fernando A. F. Ibrahim¹, Gabriel P. Carvalho¹, Luca F. Azalim¹,
Pedro H. B. de Castro¹, Renato M. A. Penna¹

¹Instituto de Ciências Exatas e Informática
Departamento de Engenharia de Software
Pontifícia Universidade Católica de Minas Gerais (PUC Minas)
Belo Horizonte – MG – Brasil

{arthur.costa¹, fafibrahim¹, gabriel.carvalho¹, lazalim¹,
pcastro¹, renato.matos¹}@sga.pucminas.br

Resumo. Este estudo investiga o perfil e o comportamento de desenvolvedores em projetos de software open source (OSS), analisando 1.000 repositórios das linguagens mais populares do GitHub. Por meio do paradigma Goal-Question-Metric (GQM), busca-se compreender como características dos contribuidores, como reputação, experiência e afiliação, impactam métricas de participação e engajamento. Os resultados indicam predominância de desenvolvedores dos Estados Unidos, Índia e China, com o Brasil em oitavo lugar. A maioria apresenta nível intermediário de experiência (88,41%), e apenas 3,74% são classificados como avançados. Contribuidores experientes alcançaram maior taxa média de aprovação de Pull Requests (83,76%) e ritmo mais constante de participação. Esses achados reforçam a influência da experiência na qualidade e regularidade das contribuições, contribuindo para o entendimento da dinâmica colaborativa em ecossistemas OSS. **Palavras-chave:** Software livre, GitHub, Experiência, Contribuidores, Engenharia de Software Empírica.

Abstract. This study investigates the profile and behavior of developers in open-source software (OSS) projects by analyzing 1,000 repositories from the most popular programming languages on GitHub. Using the Goal-Question-Metric (GQM) paradigm, it seeks to understand how contributor characteristics—such as reputation, experience, and affiliation—affect participation and engagement metrics. The results indicate a predominance of developers from the United States, India, and China, with Brazil ranking eighth. Most contributors show an intermediate experience level (88.41%), while only 3.74% are classified as advanced. Experienced contributors achieved a higher average pull request (PR) approval rate (83.76%) and a more consistent participation rhythm. These findings highlight the influence of experience on the quality and regularity of contributions, providing insights into the collaborative dynamics of OSS ecosystems. **Keywords:** Free Software, GitHub, Experience, Collaborators, Empirical Software Engineering.

1. Introdução

O desenvolvimento de software tem se transformado em uma atividade cada vez mais coletiva, apoiada por plataformas online que possibilitam a colaboração e a troca de conhecimento entre desenvolvedores [Coelho et al. 2020]. Entre essas plataformas, destaca-se o GitHub, um ambiente de hospedagem de código-fonte e controle de versão baseado em Git que permite o trabalho conjunto em projetos distribuídos, promovendo transparência, revisão contínua e construção coletiva do conhecimento técnico [Crowston et al. 2021].

Os sistemas de *open source software* (OSS) ganharam relevância não apenas pela flexibilidade tecnológica que oferecem, mas também pela força de suas comunidades [Chen and Zhou 2025]. O sucesso desses projetos depende fortemente da colaboração entre desenvolvedores distribuídos geograficamente, com diferentes níveis de experiência e motivações diversas. Nesse contexto, pesquisas recentes mostram que fatores como retenção de colaboradores, continuidade de participação e sustentabilidade comunitária são determinantes para a vitalidade dos ecossistemas [Calefato et al. 2022].

Processos como a revisão de código são centrais para garantir a qualidade do software, mas podem ser afetados por fatores externos, como reputação, experiência prévia, gênero, afiliação institucional e localização geográfica dos contribuidores. Estudos recentes indicam que fatores sociais e técnicos, como reputação, experiência prévia, gênero, afiliação institucional e localização geográfica, podem influenciar significativamente a taxa de aceitação de *pull requests* (PRs) e a percepção de confiança nas comunidades OSS (Dey e Mockus 2020).

Compreender tais processos é essencial para promover ecossistemas mais inclusivos, transparentes e eficientes [Trinkenreich et al. 2021, Dandamudi et al. 2025]. Além de contribuir para o avanço do conhecimento científico sobre a colaboração em larga escala, a análise do perfil dos contribuidores pode orientar práticas mais equitativas de gestão e engajamento comunitário [Trinkenreich et al. 2021].

Neste contexto, o objetivo deste trabalho foi caracterizar o perfil de contribuidores quanto à geolocalização, afiliação, experiência e engajamento. Para isso, conduziu-se a análise de 52743 contribuidores dos 1000 repositórios mais populares do GitHub, a fim de responder às seguintes questões de pesquisa:

- RQ1: Qual é a distribuição geográfica dos contribuidores de projetos OSS hospedados no GitHub pelo mundo?
- RQ2: Qual é a afiliação profissional dos contribuidores de projetos OSS hospedados no GitHub, e qual é o grau de dominância das empresas do grupo Meta, Apple, Nvidia, Google e Amazon (MANGA)?
- RQ3: Qual é o perfil de experiência dos contribuidores de projetos OSS hospedados no GitHub?
- RQ4: Qual é o impacto da experiência dos contribuidores na taxa pessoal de aprovação das PRs em projetos OSS hospedados no GitHub?
- RQ5: Qual a correlação da experiência dos contribuidores com a frequência de contribuição em projetos OSS hospedados no GitHub?

O restante deste artigo é organizado da seguinte forma, a Seção 2 apresenta os principais estudos correlatos. A Seção 3 descreve em detalhe a abordagem metodológica e o modelo analítico proposto. Os resultados e suas interpretações são discutidos na Seção 4,

enquanto a Seção 5 examina as ameaças à validade. Por fim, a Seção 6 encerra o artigo com as conclusões.

2. Trabalhos relacionados

Diversos estudos investigam fatores que influenciam a colaboração e a aceitação de contribuições em projetos de software de código aberto. Entre esses trabalhos, destacam-se pesquisas que abordam aspectos técnicos, sociais e demográficos dos contribuidores, bem como métricas de experiência e reputação dentro das comunidades de software OSS.

O estudo de Rastogi et al. (2018) analisou a influência da localização geográfica na avaliação de *pull requests* (PRs) em projetos do GitHub. A pesquisa identificou diferenças estatisticamente significativas nas taxas de aceitação entre os 17 países mais ativos, observando que o resultado não se correlaciona diretamente com fatores como desenvolvimento econômico ou prevalência do inglês. No entanto, o estudo encontrou que PRs submetidas por desenvolvedores residentes no mesmo país do integrador têm maior probabilidade de serem aceitas, sugerindo um viés de localização na avaliação das contribuições. Esse trabalho se relaciona ao presente estudo por também investigar se a geolocalização dos contribuidores afeta a dinâmica de participação e aceitação em projetos OSS.

Outrossim, Zhang et al. (2024) analisaram diferenças comportamentais entre desenvolvedores pagos e voluntários no ecossistema do Rust, investigando como aspectos sociais, motivacionais e profissionais influenciam o padrão de contribuições em projetos OSS. O estudo mostrou que desenvolvedores remunerados tendem a apresentar maior regularidade e constância na atividade, enquanto contribuições voluntárias estão mais associadas à motivação intrínseca e ao engajamento comunitário. Os autores também destacam que fatores como posição social dentro do projeto, histórico de participação e vínculos institucionais moldam significativamente a dinâmica de revisão, a aceitação de *pull requests* e o fluxo de colaboração. Esse trabalho é relevante por também examinar como indicadores sociais e de experiência moldam a aceitação e o engajamento dentro das comunidades de software livre.

Dey e Mockus (2020) conduziram uma análise quantitativa em larga escala sobre o ecossistema NPM (*Node Package Manager*), examinando mais de 350 mil *pull requests*. O estudo avaliou simultaneamente fatores técnicos, como tamanho e complexidade das mudanças, e fatores sociais, incluindo experiência e histórico de participação dos desenvolvedores. Os autores observaram que PRs enviadas por contribuidores mais experientes têm maior probabilidade de serem aceitas e exigem menos revisões, enquanto alterações tecnicamente mais complexas tendem a receber mais solicitações de ajustes. Esses achados reforçam que experiência acumulada e complexidade técnica influenciam diretamente a qualidade e o sucesso das contribuições em OSS.

No que se refere à caracterização da experiência dos desenvolvedores, Tamburri et al. (2019) destacam a relevância de modelos capazes de mensurar quantitativamente o conhecimento e a especialização dos contribuidores em ecossistemas colaborativos. O estudo propõe métricas baseadas em tempo de participação, diversidade de repositórios e volume de contribuições, reforçando a ideia de que a expertise pode ser inferida a partir de comportamentos observáveis nos repositórios. Neste trabalho, propõe-se o *Experience Score* (ES), uma métrica composta que sintetiza quatro dimensões da atividade do desenvolvedor: Tempo (T), Frequência (F), Volume (V) e Di-

versidade (D). O ES pondera esses fatores segundo sua relevância na literatura, atribuindo maior peso à frequência (0,45) por estar fortemente associada à expertise observável [Mockus and Herbsleb 2002, Montandon et al. 2019]. Sua formulação integra conceitos de *Experience Atoms* [Mockus and Herbsleb 2002] e *skill space* [Dey and Mockus 2020], permitindo uma avaliação escalar e comparável da experiência em projetos de código aberto.

Por fim, Calefato et al. (2022) investigaram o comportamento de contribuidores em projetos OSS, analisando o ritmo individual de participação ao longo do tempo e propondo um método para identificar períodos de inatividade por meio da variação da frequência de contribuições. Os autores demonstram que oscilações prolongadas no padrão de atividade podem indicar risco de abandono e impactar diretamente a sustentabilidade dos projetos. Essa perspectiva fundamenta a inclusão da frequência e da diversidade de atividades no ES deste trabalho, abordando o engajamento como componente essencial para compreender a dinâmica das contribuições e a experiência dos desenvolvedores.

3. Metodologia

A metodologia adotada neste estudo consiste em um *pipeline* estruturado que organiza a coleta de dados, a limpeza, o cálculo, a sumarização e a análise das informações, conforme ilustrado na Figura 1.

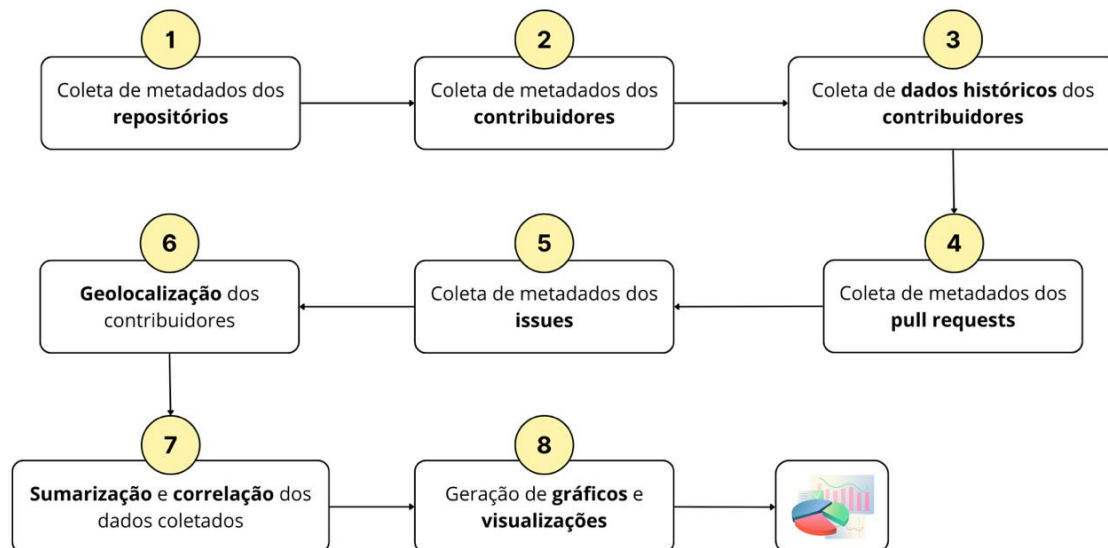


Figura 1. Representação da metodologia adotada.

3.1. Seleção dos repositórios e período de análise

A seleção dos repositórios seguiu uma metodologia sistemática baseada em dois critérios principais: diversidade de linguagens de programação e popularidade dos projetos. A plataforma GitHub foi escolhida como fonte de dados por ser amplamente reconhecida como o repositório mais popular para código aberto, sendo vista como um recurso para encontrar software de qualidade [Escamilla et al. 2022].

Com o intuito de garantir a generalidade e a representatividade dos resultados, foram consideradas as dez linguagens de programação mais populares do GitHub em 2024: *JavaScript*, *Python*, *Java*, *TypeScript*, *C#*, *C++*, *PHP*, *Shell*, *C* e *Ruby* [GitHub 2024].

Para cada linguagem, foram coletados os 100 repositórios com maior número de estrelas (`stargazerCount`), totalizando uma amostra de 1.000 repositórios. A métrica de estrelas foi adotada por representar um indicador consolidado na literatura como indicador de engajamento comunitário e atratividade de projetos OSS (Coelho et al., 2020).

O período de análise abrangeu os últimos 12 meses (**junho de 2024 a junho de 2025**) de atividade nos repositórios selecionados, considerando também o histórico completo de cada contribuidor coletado por meio da *GitHub REST* e *GraphQL API*. Essa abordagem possibilitou uma visão longitudinal da evolução das contribuições e da consistência de engajamento dos desenvolvedores.

3.2. Visão geral do *dataset*

A execução da metodologia nos 1000 repositórios resultou na identificação de 52.743 contribuidores, com a coleta de 10.014.005 PRs e 3.814.799 *issues*, representando uma base consistente para análise de perfis e histórico de contribuição.

3.3. Variáveis dependentes e independentes

Neste estudo empírico e correlacional, as variáveis independentes consistem nos fatores de perfil (geolocalização, afiliação e experiência), enquanto as variáveis dependentes correspondem aos indicadores de desempenho (taxa de aprovação de PRs e frequência de contribuição). Essa estrutura garante a rastreabilidade entre as questões de pesquisa e as métricas descritas na Tabela 1.

Tabela 1. Resumo das variáveis do estudo

Tipo de variável	Descrição
Variáveis independentes	
Geolocalização	País estimado de residência do contribuidor, obtido a partir do campo <code>location</code> do perfil público no GitHub e processado via <i>OpenStreetMap API</i> para normalização geográfica.
Afiliação	Empresa à qual o contribuidor está vinculado e se pertence ao grupo MANGA (Meta, Amazon, Netflix, Google e Apple).
Experiência (Variável Central)	<i>Experience Score</i> (ES), classificado nos níveis: Inicial (0,1–0,35), Intermediária (0,35–0,65) e Avançada (acima de 0,66).
Variáveis dependentes	
Taxa média de aprovação de PRs	Percentual de <i>Pull Requests</i> criadas por um contribuidor que foram aceitas (<i>merged</i>) pelo repositório.
Frequência de contribuição	Tempo médio em dias entre cada PR criada por cada contribuidor.

3.4. Coleta de dados:

Utilizaram-se as APIs REST e GraphQL do GitHub para extrair metadados de perfis e o histórico de PRs e *issues* dos contribuidores ativos nos últimos 12 meses. Propõe-se, neste estudo, o *Experience Score* (ES), uma métrica composta original desenvolvida para quantificar a experiência acumulada no ecossistema OSS (Mockus e Herbsleb, 2002). O ES integra as dimensões de Tempo (T), Frequência (F), Volume (V) e Diversidade (D), utilizando a fórmula $0.20T + 0.45F + 0.20V + 0.15D$. A fundamentação desta métrica apoia-se no conceito de *Experience Atoms* e na evidência de que a constância de atividade (F), que recebeu o maior peso (0,45), é o principal indicador de *expertise* observável (Montandon, Silva e Valente, 2019).

A métrica é definida pela seguinte fórmula:

$$ES = 0.20T + 0.45F + 0.20V + 0.15D$$

Experience Score

Para garantir a comparabilidade entre os diferentes componentes do score (Equação 3.4), cada métrica foi normalizada no intervalo [0, 1] antes da aplicação dos pesos definidos. A normalização foi realizada de forma min-max, considerando o valor mínimo e máximo observados em cada dimensão (Tempo, Frequência, Volume e Diversidade). Essa abordagem assegura que todas as variáveis contribuam proporcionalmente para o cálculo do ES, evitando distorções decorrentes de diferentes escalas de medição.

O escore resultante varia continuamente entre 0 e 1, permitindo comparações diretas entre contribuidores. Para fins de análise, o ES foi categorizado em três níveis de experiência (Tabela 2): Inicial, Intermediária e Avançada.

Além disso, para tornar o escore mais interpretável, o ES é categorizado em três faixas de referência. Cada intervalo foi definido para representar níveis progressivos de engajamento e maturidade técnica, distinguindo contribuidores com participação esporádica (nível inicial), atuação consistente ao longo do tempo (nível intermediário) e participação intensiva e multifacetada, típica de desenvolvedores experientes ou centrais na comunidade (nível avançado), conforme apresentado na Tabela 2.

Tabela 2. Valores de referência para o *Experience Score*

Nível de experiência	Faixa de valores
Inicial	0,10 – 0,35
Intermediária	0,35 – 0,65
Avançada	Acima de 0,66

A análise dos dados foi estruturada em duas etapas principais. Primeiramente, executou-se um processo de sumarização e correlação, gerando arquivos para cada questão de pesquisa, contendo os resultados e métricas derivadas. Em seguida, foram aplicadas técnicas estatísticas com o objetivo de verificar as hipóteses propostas. Estudos semelhantes fazem uso de regressão logística (Rastogi et al., 2018) e de testes não paramétricos, como o teste de Mann–Whitney U, para dados que não seguem distribuição normal (Zhang et al., 2024). De forma análoga, este trabalho emprega métodos de correlação e inferência estatística apropriados para relacionar o nível categórico de experiência (ES) com as métricas de aceitação e frequência de contribuição, possibilitando a identificação de padrões e relações significativas entre as variáveis analisadas.

4. Resultados

A execução do pipeline metodológico resultou em um conjunto de dados robusto proveniente de 1.000 repositórios das dez linguagens mais populares do GitHub. No total, foram processados 52.743 contribuidores ativos, 10.014.005 pull requests e 3.814.799 *issues*. Desse montante, foi possível determinar a geolocalização de 25.536 desenvolvedores (48,41%) e a afiliação profissional de 16.902 (32,04%). Esta base de larga escala fundamenta as análises apresentadas a seguir, organizadas conforme as cinco questões de pesquisa propostas.

4.1. Q1: Como os contribuidores de projetos OSS do GitHub estão distribuídos geograficamente pelo mundo?

A Figura 2 apresenta os nove países com maior número de contribuidores identificados, oferecendo uma visão geral da distribuição geográfica dos participantes.

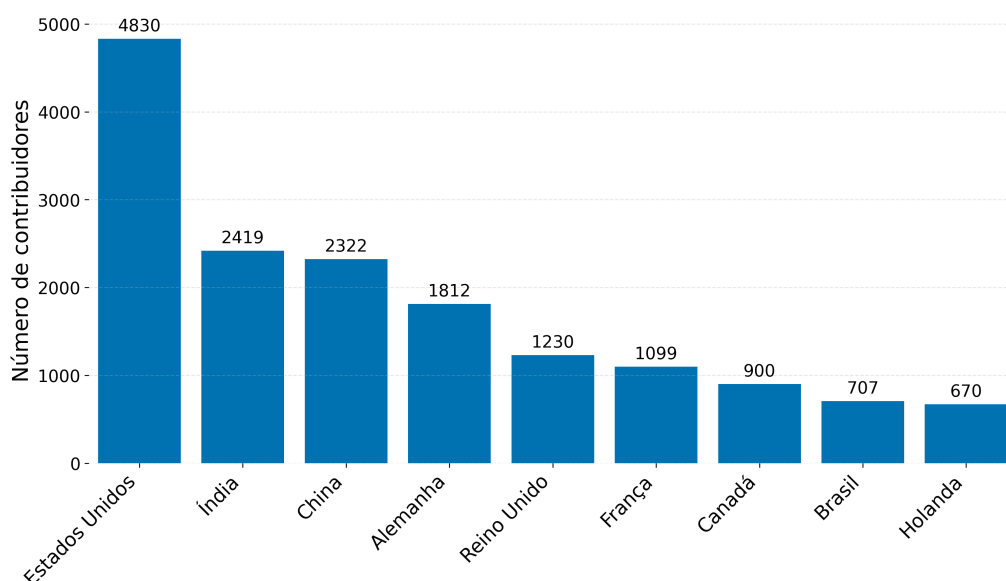


Figura 2. Distribuição geográfica dos contribuidores de projetos OSS.

A análise da distribuição geográfica identificou um total de 52.743 contribuidores, dos quais 25.536 tiveram a localização determinada, representando uma cobertura de 48,41%. Esses resultados evidenciam uma predominância de participantes oriundos da América do Norte e da Ásia, refletindo a concentração global de polos tecnológicos e de comunidades OSS nessas regiões.

4.2. Q2: Qual é a afiliação profissional dos contribuidores de projetos OSS no GitHub, e qual é o grau de dominância das empresas do grupo MANGA?

A identificação de vínculos institucionais é crucial para entender a sustentabilidade de projetos OSS, uma vez que desenvolvedores pagos tendem a apresentar maior regularidade de contribuição [Zhang et al. 2024]. Nesta análise (Q2), determinou-se a afiliação de 16.902 contribuidores (32,04% da base), sendo que o restante (67,96%) não utilizou identificadores corporativos. Este grupo majoritário é caracterizado como contribuidores independentes ou voluntários, que participam motivados por interesses pessoais ou profissionais não explicitados em seus perfis

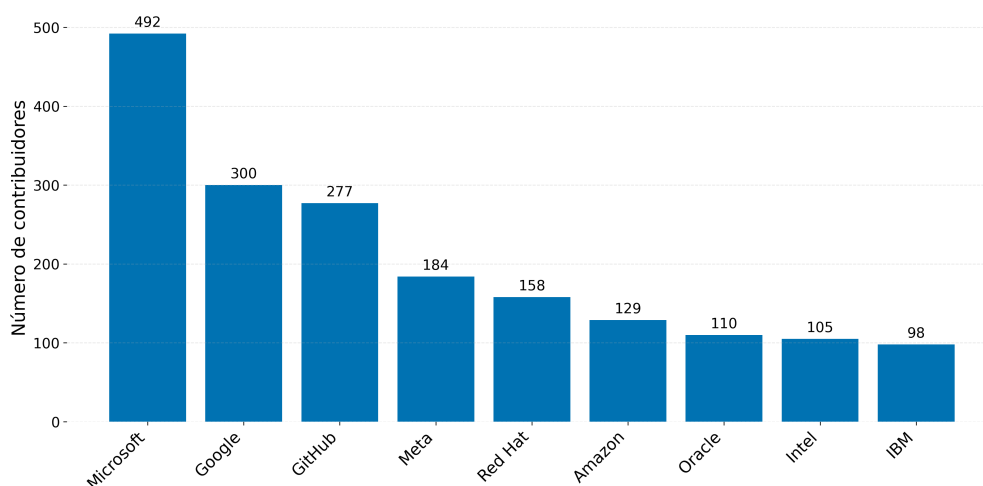


Figura 3. Empresas com maior número de contribuidores identificados.

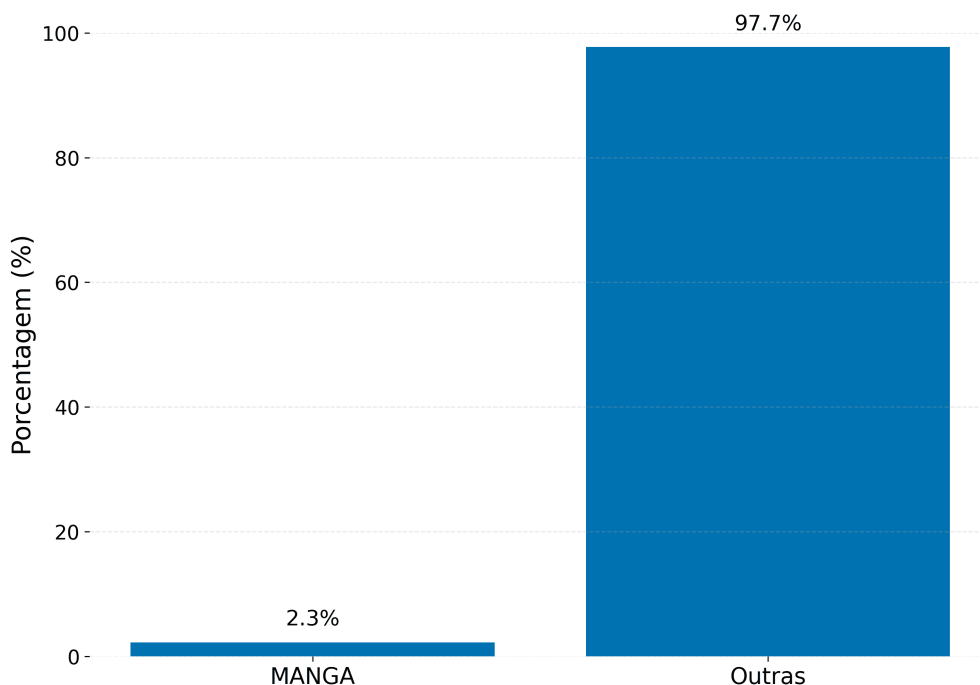


Figura 4. Comparação entre empresas MANGA vs Outras.

Os resultados (Figura 3 e Figure 4) revelam que, embora gigantes tecnológicas como Microsoft (492) e Google (300) liderem as contribuições, o ecossistema é amplamente descentralizado. O grupo MANGA representou apenas 2,3% do total analisado, indicando que a dinâmica colaborativa é sustentada predominantemente por esforços comunitários e organizações de menor escala, e não apenas por grandes blocos corporativos específicos

4.3. Q3: Qual é o perfil de experiência dos contribuidores de projetos OSS no GitHub?

Para complementar a análise da distribuição dos níveis de experiência, a Figura 5 apresenta um histograma do ES calculado para todos os contribuidores. Essa visualização permite observar a forma da distribuição contínua do escore, destacando sua concentração em valores intermediários e evidenciando a assimetria que caracteriza o perfil predominante da comunidade analisada.

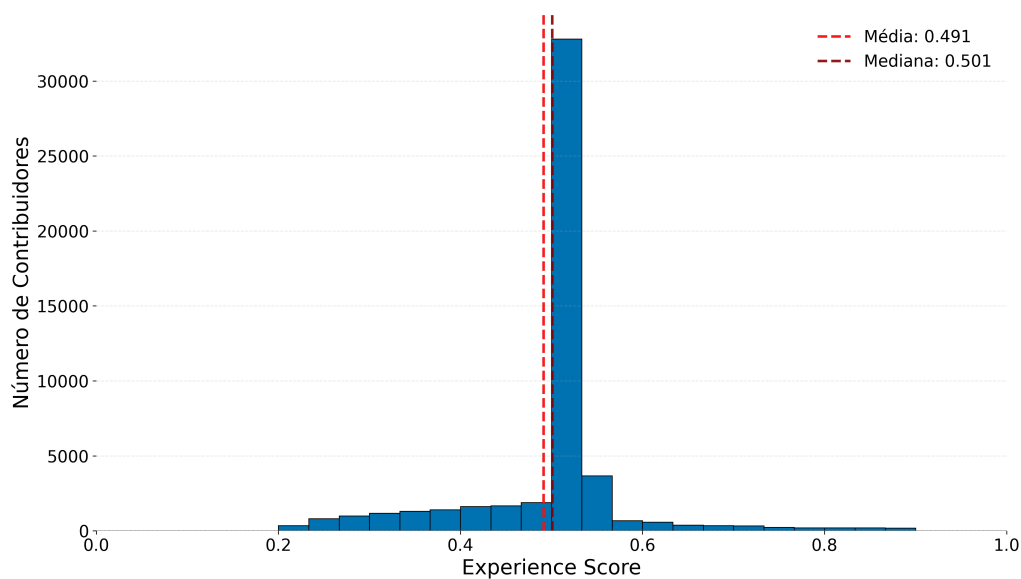


Figura 5. Distribuição dos níveis de experiência dos contribuidores.

O *Experience Score* (ES) foi calculado para 52.743 contribuidores, considerando pesos de 0,45 para regularidade, 0,20 para duração, 0,20 para volume e 0,15 para diversidade. A média do escore foi de 0,491, com mediana de 0,501. Representado na Figura 5, a maioria pertence ao nível Intermediário (88,41%), seguida dos níveis Inicial (7,85%) e Avançado (3,74%).

Esses resultados indicam que a maior parte dos desenvolvedores OSS possui envolvimento consistente, porém ainda limitado em termos de amplitude e profundidade técnica, caracterizando um perfil predominantemente intermediário. Dessa maneira, pode-se inferir que o ecossistema OSS é composto, em sua maioria, por desenvolvedores com experiência prática consolidada, mas ainda em processo de especialização. Isso reforça a importância das comunidades OSS como espaços de aprendizado contínuo e de formação de novos especialistas.

4.4. Q4: Como a experiência dos contribuidores afeta a taxa pessoal de aprovação das PRs em projetos OSS no GitHub?

A Figura 6 apresenta a relação entre o nível de experiência e a taxa de aprovação de *pull requests*, evidenciando o impacto direto do ES no processo de revisão. Os resultados mostram que contribuidores classificados no nível Avançado apresentaram taxa média de aceitação de 83,76%, enquanto os níveis Intermediário e Inicial registraram, respectivamente, 57,74% e 56,24%. Essa diferença expressiva indica que contribuintes com maior experiência tendem a submeter PRs mais alinhadas às normas dos projetos e com menor necessidade de retrabalho.

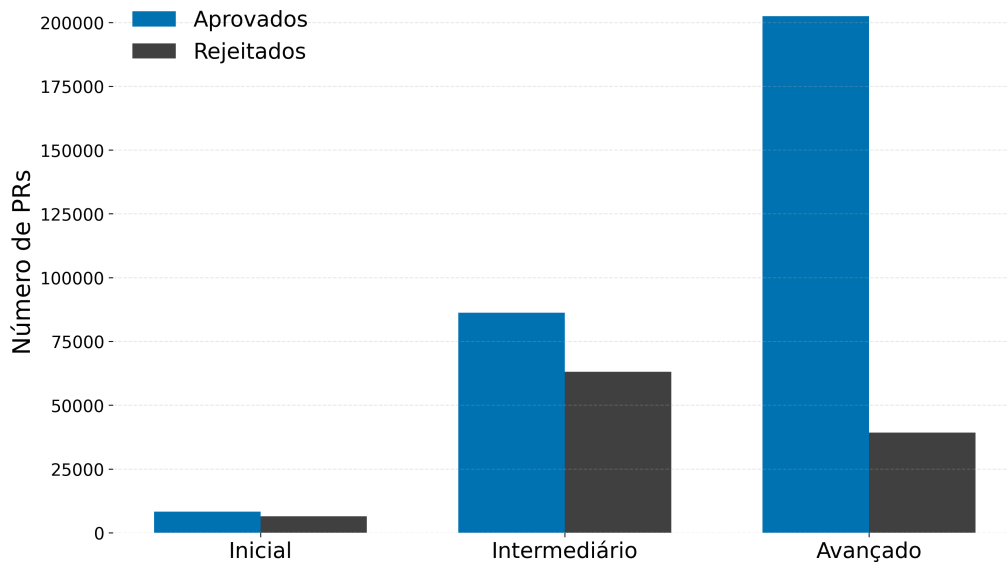


Figura 6. Tendência de aprovação de PRs em função do Experience Score.

Assim, observa-se uma correlação positiva entre experiência e aceitação das contribuições. Esses achados permitem rejeitar a hipótese nula (H_0) estabelecida para a Q4 e aceitar a hipótese alternativa (H_A), concluindo que há uma relação estatisticamente significativa entre o nível de experiência e a taxa de aprovação das PRs. Pode-se inferir que a experiência acumulada pode atuar, portanto, como fator determinante para o reconhecimento técnico e a confiabilidade dos contribuidores no processo de revisão de código.

4.5. Q5: Como a experiência dos contribuidores está correlacionada à frequência de contribuição em projetos OSS no GitHub?

A Figura 7 apresenta a relação entre o nível de experiência dos contribuidores e a frequência com que submetem *pull requests*, permitindo observar como o ritmo individual de participação varia entre os diferentes perfis de experiência.

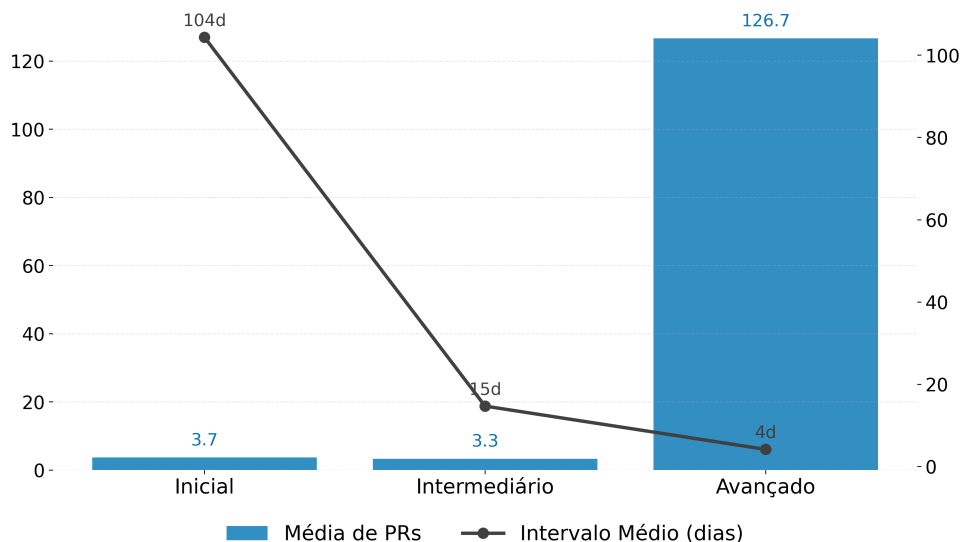


Figura 7. Intervalo médio entre PRs por nível de experiência.

Os resultados apresentados na Figura 7 demonstram que contribuidores do nível Avançado registraram média de 126,67 PRs por contribuidor, com intervalo médio de 4,14 dias entre submissões. Em contraste, os níveis Intermediário e Inicial apresentaram médias de 3,31 PRs (intervalo de 14,66 dias) e 3,65 PRs (intervalo de 104,33 dias), respectivamente.

Esses dados indicam uma correlação positiva entre o *Experience Score* e a frequência de contribuições, demonstrando que desenvolvedores mais experientes mantêm um ritmo mais constante e engajado de participação. Dessa forma, rejeita-se a hipótese nula (H_0) e confirma-se a hipótese alternativa (H_A) para a Q5, uma vez que o aumento da experiência está associado a maior constância e produtividade. Em conjunto com os resultados da Q4, este achado reforça a importância da regularidade e da prática contínua como elementos centrais para a consolidação da expertise em ecossistemas OSS.

5. Discussão

Os resultados obtidos neste estudo revelam padrões consistentes sobre o perfil e o comportamento de contribuidores em projetos OSS, permitindo uma análise integrada das dimensões geográficas, profissionais e de experiência. Ao observar conjuntamente as questões de pesquisa, emergiram evidências que ajudam a compreender como fatores individuais influenciam tanto a dinâmica de colaboração quanto o desempenho técnico dentro dos ecossistemas OSS.

Primeiramente, a distribuição geográfica (Q1) confirma a predominância de contribuidores em regiões com maior concentração tecnológica, como Estados Unidos, Índia e China. Esse dado, embora esperado, reforça a persistência de assimetrias globais na participação em projetos OSS. A presença significativa de países do hemisfério norte pode indicar que oportunidades profissionais e proficiência em inglês continuam a desempenhar papel central na participação colaborativa. O Brasil, embora apareça entre os dez países mais ativos, ainda apresenta representatividade relativamente reduzida, evidenciando desafios de inclusão geográfica.

A análise de afiliação (Q2) revelou uma comunidade amplamente descentralizada, convergindo com os achados de geolocalização (Q1) ao demonstrar que o ecossistema OSS não foi dominado por poucos centros de poder técnico ou geográfico. Embora grandes corporações globais tenham apresentado participação ativa, elas representaram apenas uma fração mínima (2,3%) do total de contribuidores analisados, o que evidenciou que o desenvolvimento foi sustentado majoritariamente por indivíduos independentes e organizações de menor escala. Essa pluralidade institucional, distribuída por diversas regiões e polos tecnológicos, reforçou o caráter comunitário da colaboração em larga escala, sinalizando que a dinâmica do software aberto transcende a influência de grandes conglomerados tecnológicos específicos e se apoia na diversidade de contribuições locais e voluntárias.

No que diz respeito à experiência (Q3), a forte concentração de contribuidores no nível Intermediário (88,41%) tem implicações importantes. Esse achado demonstra que o OSS funciona como um ambiente de aprendizagem contínua, onde muitos desenvolvedores mantêm atividade moderada, mas ainda não atingem níveis elevados de especialização. O pequeno percentual de contribuidores Avançados (3,74%) evidencia um fenômeno já documentado na literatura: poucas pessoas respondem por grande parte das contribuições

regulares, atuando como líderes técnicos e pilares de sustentabilidade dos projetos.

As questões (Q4 e Q5) reforçam o papel central da experiência no desempenho técnico. Contribuidores mais experientes apresentaram taxas de aprovação de PRs significativamente mais altas, além de maior constância no envio de contribuições. Esses achados sugerem que a experiência não apenas aprimora a qualidade das contribuições, mas também facilita a compreensão das normas e padrões implícitos de cada projeto, reduzindo atritos no processo de revisão. A regularidade observada em contribuidores avançados aponta para um ciclo virtuoso entre prática frequente e desenvolvimento de expertise.

Ao integrar esses resultados, observa-se que a dinâmica de participação em OSS é influenciada simultaneamente por fatores estruturais (como localização e afiliação) e individuais (como experiência e engajamento). Enquanto fatores geográficos e profissionais ajudam a contextualizar o ambiente de participação, é a experiência — medida pelo *Experience Score* — que mais fortemente se conecta a métricas de desempenho. Assim, a experiência emerge como o principal indicador de maturidade técnica e social dentro das comunidades OSS, influenciando tanto a aceitação das contribuições quanto o ritmo de participação.

Por fim, esses resultados dialogam com trabalhos recentes que destacam a importância da reputação, do histórico de participação e da constância como elementos que moldam a colaboração em larga escala. A utilização de métricas compostas como o *Experience Score* mostra-se promissora para estudos futuros, podendo apoiar mantenedores na identificação de contribuidores especialistas, antecipação de riscos de abandono e melhoria dos fluxos de triagem e revisão.

6. Ameaças à validade

A principal ameaça à validade de constructo reside na mensuração da experiência por meio do *Experience Score* (ES). Embora o ES utilize atividades observáveis como *proxy* para a senioridade, essa métrica é um indicador parcial que pode não capturar dimensões cognitivas da *expertise*, como habilidades de design e comunicação. Adicionalmente, o foco predominante em contribuições de código (*commits* e PRs) omite outras formas essenciais de engajamento, como a revisão de código por pares e discussões técnicas, que são fundamentais para a dinâmica colaborativa. Em relação às validades interna e externa, o estudo não controlou fatores intervenientes que influenciam o sucesso das contribuições, tais como a complexidade intrínseca do código submetido, a disponibilidade de revisores e a natureza do vínculo (pago ou voluntário) do desenvolvedor. Por fim, a generalização das conclusões é limitada ao ecossistema de repositórios públicos do GitHub e ao recorte temporal de 12 meses. Resultados plenamente generalizáveis para o ecossistema OSS global exigiriam a realização de complexos experimentos que replicassem esta metodologia em outras plataformas e contextos tecnológicos distintos.

7. Conclusão

Este estudo investigou a relação entre a experiência de desenvolvedores e seus padrões de colaboração e sucesso em projetos de software *open source* (OSS) hospedados no GitHub. A partir de uma abordagem empírica quantitativa baseada na mineração de repositórios, foi proposto e aplicado o ES, uma métrica composta que quantifica a experiência de um

contribuidor com base na ponderação de frequência, volume, tempo e diversidade de suas atividades.

Os resultados demonstram que o nível de experiência exerce influência significativa sobre o desempenho e a integração social dos contribuidores. Contribuidores classificados com ES elevado apresentaram taxas médias de aceitação de PRs superiores, maior regularidade de participação e um padrão de engajamento mais constante ao longo do tempo.

Esses achados reforçam a hipótese de que a experiência funciona como mecanismo social e técnico central no ecossistema OSS, impactando diretamente a dinâmica de colaboração, a produtividade e a eficiência dos processos de revisão de código. Em particular, observou-se que desenvolvedores mais experientes tendem a receber respostas mais rápidas, ter maior aceitação de suas contribuições e exercer papel de liderança técnica nos projetos.

Em síntese, a capacidade de quantificar a experiência por meio do ES oferece um sinal objetivo que pode ser explorado por mantenedores de projetos OSS e plataformas de colaboração, tanto para fins de reconhecimento de contribuições quanto para a melhoria de processos de triagem e revisão de código.

Como continuidade deste trabalho, propõe-se expandir o escopo da pesquisa para outras plataformas além do GitHub, incorporar métricas sociais e qualitativas que capturem interações entre desenvolvedores e aprimorar o *Experience Score* com novas dimensões que reflitam melhor os aspectos técnicos e sociais da experiência dos contribuidores.

Referências

- Calefato, F., Gerosa, M. A., Iaffaldano, G., Lanubile, F., and Steinmacher, I. (2022). Will you come back to contribute? investigating the inactivity of oss core developers in github. *Empirical Software Engineering*, 27(3).
- Chen, X. and Zhou, Y. (2025). Open-source collaboration and technological innovation in the industrial software industry: A multi-case study. *Systems*, 13(6).
- Coelho, J., Valente, M. T., Milen, L., and Silva, L. L. (2020). Is this github project maintained? measuring the level of maintenance activity of open-source projects. *Information and Software Technology*, 122:106274.
- Creswell, J. W. (2010). *Projeto de pesquisa: métodos qualitativo, quantitativo e misto*. Bookman, Porto Alegre, 3 edition.
- Crowston, K., Shamshurin, I., and Bolici, F. (2021). Coordination and code review in distributed open source teams. *Information and Software Technology*, 138:106636.
- Dandamudi, R., Adaji, I., and Rodríguez-Pérez, G. (2025). Reflection on code contributor demographics and collaboration patterns in the rust community. In *Proceedings of The 29th International Conference on Evaluation and Assessment in Software Engineering (EASE 2025)*. ACM.
- Dey, T., Karnauch, A., and Mockus, A. (2020). Representation of developer expertise in open source software. In *ACM Conference*, page 11.

- Dey, T. and Mockus, A. (2020). Effect of technical and social factors on pull request quality for the npm ecosystem. *arXiv preprint arXiv:2007.04816*.
- Escamilla, J., Klein, M., Cooper, D., Rampin, R., Weigle, M. C., and Nelson, M. L. (2022). The rise of github in scholarly publications. *arXiv preprint arXiv:2208.04895*.
- GitHub (2024). The 2024 state of the octoverse: How open source is fueling the AI era. Acessado em: 21 abr. 2026.
- Lima, A., Rossi, L., and Musolesi, M. (2014). Coding together at scale: Github as a collaborative social network. In *Proceedings of the [Conference Name]*, Birmingham, UK (Instituição de afiliação dos autores). Association for the Advancement of Artificial Intelligence. Copyright ©2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved. [2].
- Lin, B., Robles, G., and Serebrenik, A. (2017). Developer turnover in global, industrial open source projects: Insights from applying survival analysis. In *IEEE 12th International Conference on Global Software Engineering*, pages 66–75.
- Mockus, A. and Herbsleb, J. D. (2002). Expertise browser: A quantitative approach to identifying expertise. In *International Conference on Software Engineering (ICSE)*, pages 503–512, Orlando, Florida.
- Montandon, J. E., Silva, L. L., and Valente, M. T. (2019). Identifying experts in software libraries and frameworks among github users. In *IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*, pages 276–287.
- Perez, Q., Urtado, C., and Vauttier, S. (2023). Dataset of open-source software developers labeled by their experience level in the project and their associated software metrics. *Data in Brief*, 46:108842.
- Rastogi, A., Nagappan, N., Gousios, G., and Hoek, A. v. d. (2018). Relationship between geographical location and evaluation of developer contributions in github. In *ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, page 8, Oulu, Finland. ACM.
- Trinkenreich, B., Wiese, I., Sarma, A., Gerosa, M., and Steinmacher, I. (2021). Women’s participation in open source software: A survey of the literature. *ACM Transactions on Software Engineering and Methodology*, 1(1):36.
- Wazlawick, R. S. (2014). *Metodologia de pesquisa para ciência da computação*. Elsevier, Rio de Janeiro, 2 edition.
- Zhang, Y., Qin, M., Stol, K.-J., Zhou, M., and Liu, H. (2024). How are paid and volunteer open source developers different? a study of the rust project. In *2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE ’24)*, page 13.