

## **Investigando o Envolvimento de Empresas em Projetos de Software Aberto**

## **Investigating How Companies Get Involved in Open Source Software Projects**

**Jonas F. de Souza<sup>1</sup>, Bruno Cartaxo<sup>1</sup>**

<sup>1</sup> Análise e Desenvolvimento de Sistemas -- Instituto Federal de Pernambuco (IFPE)  
Paulista – PE – Brasil

[jonasfelixdesouza21@gmail.com](mailto:jonasfelixdesouza21@gmail.com) , [email@brunocartaxo.com](mailto:email@brunocartaxo.com)

---

**Resumo.** É amplamente reconhecido que voluntários e indivíduos altruístas são os principais impulsionadores da maioria dos projetos de código aberto (OSS - Open Source Software). No entanto, nos últimos anos, essa dinâmica tem mudado à medida que a indústria de desenvolvimento de software se torna cada vez mais dependente desses projetos. Nesse cenário, este artigo investiga como as empresas têm apoiado projetos de código aberto. Para isso, analisamos um conjunto de dados com 17.264 projetos de OSS apoiados por empresas disponíveis no GitHub.

Nossos resultados revelam que um número significativo de empresas (> 4000) está se envolvendo com OSS, embora raramente adotem projetos preexistentes, preferindo apoiar aqueles que iniciaram ou que possuem. Além disso, os projetos OSS mais comuns apoiados por empresas concentram-se em bibliotecas, frameworks e ferramentas, enquanto o código de suas aplicações, frequentemente considerado uma vantagem competitiva, raramente é disponibilizado.

**Palavras-chave:** OSS, Software de Código Aberto, Software, Apoio de Empresas, GitHub

**Abstract.** It is widely recognized that volunteers and altruistic individuals are the main drivers of most Open Source Software (OSS) projects. However, in recent years, this dynamic has been shifting as the software development industry becomes increasingly dependent on these projects. In this context, this paper examines how companies are supporting OSS projects. To achieve this goal, we analyzed a dataset of 17,264 company-supported OSS projects available on GitHub.

Our findings reveal that a significant number of companies (> 4000) are engaging with OSS; however, they rarely adopt preexisting OSS projects, preferring instead to support those they have initiated or own. Furthermore, the most common domains of company-supported OSS projects include libraries, frameworks, and tools, while the source code of their applications, often considered a competitive advantage, is rarely made available.

**Key words:** OSS, Open Source Software, Software, Company Support, GitHub

---

# 1. Introdução

Quando o movimento de Software de Código Aberto (OSS) começou, a maioria dos projetos era mantida pelo esforço voluntário de indivíduos. No entanto, esse cenário mudou à medida que a indústria de desenvolvimento de software passou a depender fortemente do OSS (Pinto et al., 2018).

Dentre as múltiplas soluções OSS utilizadas na grande maioria dos sistemas de software modernos, incluem-se: ferramentas de construção e teste, IDEs, servidores web, frameworks web, bancos de dados e muito mais. Dessa forma, empresas e organizações começaram a se envolver em projetos OSS, incluindo Microsoft, Google, Amazon, Apple e Meta. Para ilustrar esse fenômeno, mais de 80% das contribuições para o kernel Linux são feitas por desenvolvedores que são pagos para isso (Foundation, 2015).

Um estudo de 2018 investigou o repositório de código de cinco projetos OSS e observou que mais da metade das contribuições para esses projetos veio de desenvolvedores pagos por empresas (Dias, Steinmacher, Pinto, 2018). Outra pesquisa publicada recentemente reuniu um conjunto de dados com mais de 17.000 projetos OSS disponíveis no GitHub que são desenvolvidos principalmente por empresas, e não por voluntários (Spinellis et al., 2020).

Compreender esse fenômeno pode ajudar a comunidade OSS a aproveitar melhor as parcerias com empresas, bem como auxiliar empresas que desejam contribuir para projetos OSS. Assim, o objetivo desta pesquisa é:

Investigar como empresas e organizações se envolvem com projetos OSS.

## 2. Trabalhos Relacionados

A literatura sobre o envolvimento de empresas privadas em projetos de software de código aberto (OSS) é extensa e diversa. Um estudo sistemático de 92 pesquisas relevantes (Li et al., 2024) examinou as motivações por trás do envolvimento corporativo no OSS, os diversos modelos de colaboração entre organizações e comunidades de desenvolvedores, e o impacto mais amplo dessas interações nos negócios, projetos e no ecossistema OSS.

Uma constatação importante desse estudo é que as análises quantitativas de repositórios de software ainda são relativamente escassas. Quando tais análises são realizadas, elas geralmente focam em um único repositório ou em um pequeno conjunto de projetos.

Com base nisso, diversos estudos de caso buscaram preencher essa lacuna, investigando repositórios e ecossistemas específicos de OSS. Zhang et al. (Zhang et al., 2020), por exemplo, analisaram o ecossistema OpenStack, examinando históricos de commits para identificar padrões de colaboração entre as empresas participantes. Da mesma forma, Dias et al. I. (Dias; Steinmacher; Pinto, 2018) estudaram repositórios de projetos como Atom, Electron, Git-LFS, Hubot e Linguist, todos hospedados no GitHub.

O estudo comparou contribuições de desenvolvedores externos e funcionários internos (da GitHub), usando históricos de commits e outros metadados. Os resultados revelaram que, embora os

desenvolvedores internos fossem menores em número, eles contribuíam com mais frequência. Por outro lado, os desenvolvedores externos desempenharam um papel crucial em tarefas complexas e no desenvolvimento geral do código, sendo responsáveis por 56% dos pull requests.

Mais recentemente, Zhang et al. (Zhang et al., 2024) analisaram o repositório do Rust, investigando diferenças entre desenvolvedores pagos e voluntários. Além da análise de commits, os autores entrevistaram colaboradores do projeto para entender as percepções dos voluntários sobre os desenvolvedores pagos.

Osborne et al. (Osborne et al., 2024) investigaram padrões de coopetição entre empresas que contribuem para PyTorch, TensorFlow e Transformers. Utilizando mineração de dados de commits e outros métodos analíticos, examinaram como organizações concorrentes colaboram nesses repositórios. Seus achados revelam que, embora a coopetição ocorra, a empresa proprietária do repositório é responsável por até 80% dos commits, destacando seu papel dominante no desenvolvimento.

Embora esses estudos ofereçam insights valiosos, seu escopo é frequentemente limitado, geralmente focando em ecossistemas individuais ou pequenos grupos de repositórios. Em contraste, nossa pesquisa adota uma abordagem mais ampla, analisando um conjunto significativamente maior e mais diverso de software de código aberto. Ao aproveitar esse extenso conjunto de dados, buscamos fornecer uma compreensão mais abrangente da participação corporativa no OSS, revelando tendências e padrões que podem não surgir em estudos de menor escala.

Além disso, nosso trabalho introduz uma contribuição inédita ao explorar os domínios específicos dos projetos nos quais as empresas geralmente se envolvem, uma área ainda não abordada na literatura existente.

### 3. Metodologia

Aqui, apresentamos nossas perguntas de pesquisa (Seção 3.1) e a justificativa por trás delas, bem como a forma como organizamos os dados (Seção 3.2). Além disso, também disponibilizamos um pacote para apoiar qualquer pessoa que queira replicar este estudo<sup>1</sup>.

#### 3.1. Perguntas de Pesquisa

Para alcançar o objetivo enunciado na Seção 1, buscamos responder às seguintes perguntas de pesquisa:

RQ1: Quantas e quais empresas contribuem para projetos OSS?

**Justificativa:** Ao responder a essa pergunta, podemos ter uma ideia se há muitas ou poucas empresas investindo em projetos OSS e também se algumas empresas concentram a maior parte das contribuições.

RQ2: Com que frequência as empresas adotam projetos preexistentes?

---

<sup>1</sup> <https://doi.org/10.5281/zenodo.7553532>

**Justificativa:** Com essa questão, buscamos descobrir se as empresas se envolvem apenas nos projetos que iniciam e possuem ou se adotam projetos preexistentes. Isso pode revelar como as empresas se relacionam com a iniciativa OSS como um todo.

RQ3: Quais são os domínios desses projetos?

**Justificativa:** Ao responder a essa pergunta, podemos entender os tipos de projetos OSS nos quais as empresas geralmente investem. Para respondê-la, adotamos a classificação de domínios proposta por Borges et al. (Borges; Hora; Valente, 2016), que inclui:

- Software de Aplicação: sistemas que fornecem funcionalidades para usuários finais, como navegadores e editores de texto;
- Software de Sistema: sistemas que fornecem serviços e infraestrutura para outros sistemas, como sistemas operacionais, middlewares, servidores e bancos de dados;
- Bibliotecas e Frameworks Web: como Bootstrap, Angular e React;
- Bibliotecas e Frameworks Não-Web: como Google Guava, Facebook Fresco, entre outros;
- Ferramentas de Software: sistemas que suportam tarefas de desenvolvimento de software, como IDEs, gerenciadores de pacotes, compiladores e sistemas de controle de versão;
- Documentação: repositórios com documentação, tutoriais e exemplos de código-fonte.

RQ4: Como os commits de contribuidores de empresas e não empresas variam ao longo da semana?

**Justificativa:** Com essa questão, buscamos entender se há diferença no padrão de contribuição — especificamente, no número de commits feitos em cada dia da semana — entre funcionários pagos por empresas e contribuidores eventuais/hobbyistas de projetos OSS.

RQ5: Quantos e quais projetos são apoiados por mais de uma empresa?

**Justificativa:** Com essa questão, queremos identificar se as empresas apoiam projetos que também são suportados por outras empresas ou se tendem a investir mais em projetos nos quais ditam os próximos passos sozinhas.

## 3.2. Curadoria de Dados

Esta pesquisa se baseia principalmente no conjunto de dados de Spinellis et al. (Spinellis et al., 2020)<sup>2</sup>, que fornece informações sobre 17.252 repositórios do GitHub, incluindo a empresa proprietária do projeto no GitHub e o domínio de e-mail dominante dos projetos, entre outros.

No entanto, **para responder à RQ1**, tivemos que lidar com dados ausentes, em particular 13.086 repositórios sem um nome de empresa associado. Para preencher essa lacuna, seguimos três etapas:

1. **Consulta Whois:** Criamos um script para usar o domínio de e-mail dominante de cada repositório para procurar informações em um servidor Whois, o que nos permitiu identificar automaticamente 7.865 nomes de empresas ausentes. Isso deixou 5.221 repositórios sem nome de empresa devido a metadados Whois privados, que não são acessíveis.
2. **Consulta à API do GitHub:** Criamos um script para extrair informações sobre os usuários do GitHub que são proprietários dos repositórios, especificamente seus

---

<sup>2</sup> <https://doi.org/10.5281/zenodo.3653877>

nomes de usuário, login, empresa e nome de usuário no Twitter. Essa abordagem nos ajudou a identificar mais dados ausentes.

3. **Análise manual:** Analisamos manualmente as informações retornadas pela API do GitHub e conseguimos determinar os nomes das empresas restantes.

**Para responder à RQ2**, utilizamos a API do GitHub para verificar quantos repositórios do conjunto de dados são forks. Se uma empresa faz fork de um projeto independente, podemos dizer que ela adotou esse projeto.

Para evitar perder forks informais, quando uma empresa copia o código de um repositório preexistente e cria um novo sem usar a funcionalidade de fork do GitHub, criamos um script que pesquisou o termo "fork" nos arquivos README e na descrição dos repositórios. Posteriormente, analisamos manualmente os textos completos para identificar se eram de fato forks informais ou apenas mencionavam a palavra "fork" em outro contexto.

**Para responder à RQ3**, classificamos automaticamente os repositórios do conjunto de dados de Spinellis et al. (Spinellis et al., 2020) ainda disponíveis no GitHub em setembro de 2022 (16.371 dos 17.252) utilizando o modelo de aprendizado de máquina criado por Zanartu et al. (Zanartu et al., 2022). Esse modelo foi desenvolvido com base no conjunto de dados anotado por Borges et al. (Borges; Hora; Valente, 2016), que inclui 5.000 repositórios do GitHub classificados por domínio.

Para obter melhor precisão, o modelo de Zanartu et al. combinou as categorias "Software de Aplicação" e "Software de Sistema", resultando em cinco categorias em vez das seis originalmente propostas por Borges et al.

**Para responder à RQ4**, todos os repositórios ainda disponíveis online e presentes no conjunto de dados foram clonados, minerados e tiveram seus dados analisados. Extraímos a data e a hora dos commits em todos os repositórios. Com esses dados, foi possível determinar quantos commits foram feitos em cada dia da semana.

Para refinar os resultados, o script criado para responder a essa questão gerou dois conjuntos de dados, um com commits de todos os domínios de e-mail e outro com um filtro para excluir domínios de e-mail comuns, como Gmail e Hotmail. A ideia aqui é que commits feitos com domínios de e-mail comuns geralmente são realizados por contribuidores casuais/eventuais (Pinto; Steinmacher; Gerosa, 2016/9/Saner.2016.68), enquanto aqueles com domínios mais especializados tendem a representar atividade corporativa.

**Para responder à RQ5**, expandimos a coluna de domínio de e-mail dominante do conjunto de dados de Spinellis et al. para incluir os cinco principais domínios de e-mail dominantes. Mantivemos a lógica de que cada domínio de e-mail diferente representa uma empresa distinta contribuindo para um repositório. Também ignoramos domínios de e-mail comuns como gmail.com, outlook.com, yahoo.com, entre outros, presentes em uma lista pública<sup>3</sup>

## 4. Resultados e Discussões

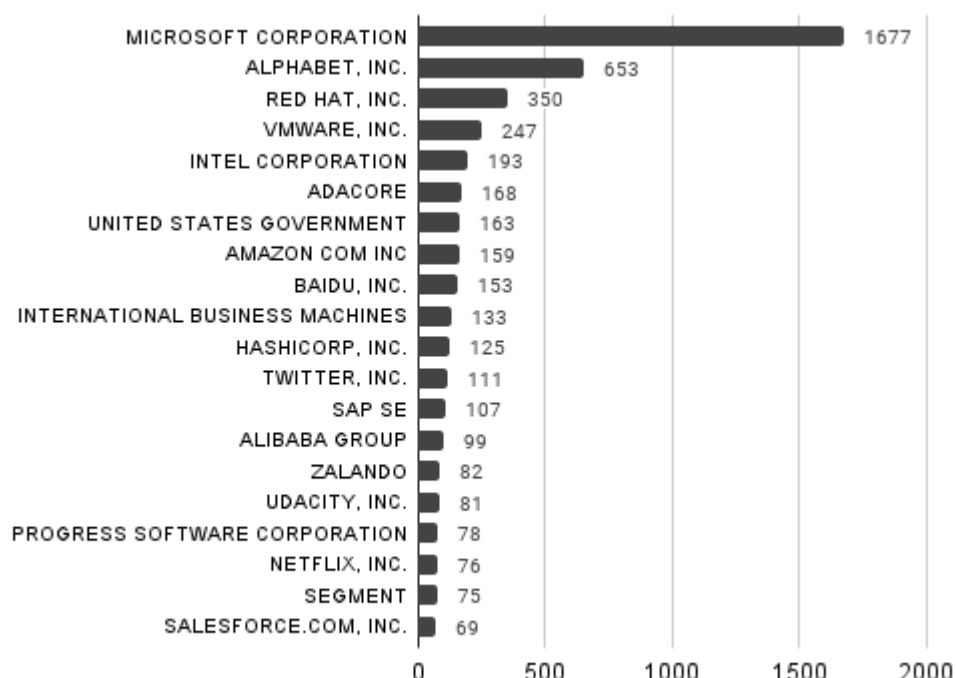
Nesta seção, apresentamos as evidências que sustentam as respostas para cada uma das perguntas de pesquisa.

---

<sup>3</sup> <https://gist.github.com/okutbay/5b4974b70673dfdcc21c517632c1f984>

## 4.1. Quantas e quais empresas contribuem para projetos OSS?

Identificamos 4.060 empresas únicas que apoiam projetos OSS. A Fig.1 mostra as 20 principais empresas que mais contribuem para projetos OSS.



**Figura 1. As 20 principais empresas que mais contribuem para projetos OSS.**

Como podemos ver, a Microsoft é a empresa com o maior número de projetos nos quais é a principal contribuidora, apoiando mais de duas vezes o número de projetos OSS (total de 1.677) em comparação com a segunda colocada, Alphabet, que apoia 653 projetos OSS. Esses números evidenciam a mudança cultural da Microsoft em direção ao software de código aberto nos últimos anos.

Por outro lado, nem sempre as empresas contribuem para um grande número de projetos OSS. O número de empresas que contribuem para apenas um projeto é alto: 2.531 empresas, ou aproximadamente 60% das empresas do conjunto de dados.

Vale destacar que entre as 10 principais, temos o Governo dos Estados Unidos, colaborando com 163 projetos OSS por meio de diversas instituições governamentais agrupadas sob o título de "Governo dos Estados Unidos". Portanto, não encontramos apenas a presença de empresas privadas, mas também de instituições públicas.

Além disso, observamos a predominância de organizações dos EUA entre as 20 principais, com uma exceção notável: o Alibaba Group, uma empresa chinesa.

## 4.2. Com que frequência as empresas adotam projetos preexistentes?

Ao consultarmos a API do GitHub para identificar repositórios marcados como forks, encontramos apenas 92 entre os mais de 17 mil do conjunto de dados. Após uma análise manual detalhada,

descobrimos que a maioria desses forks (87) são, na verdade, "forks internos", ou seja, uma cópia pessoal do projeto feita por um funcionário da empresa ou uma mudança do repositório onde a equipe de desenvolvimento submete suas contribuições.

Assim, apenas 5 dos 92 projetos marcados como forks foram de fato adotados por empresas. Como esse número é insignificante em relação ao total de repositórios, buscamos por "forks informais".

Depois de analisar manualmente os textos, identificamos apenas 3 repositórios que se enquadram na definição de fork informal. Em resumo, apenas 8 dos mais de 17 mil projetos OSS do conjunto de dados são forks reais, cinco usando a funcionalidade do GitHub e três forks informais.

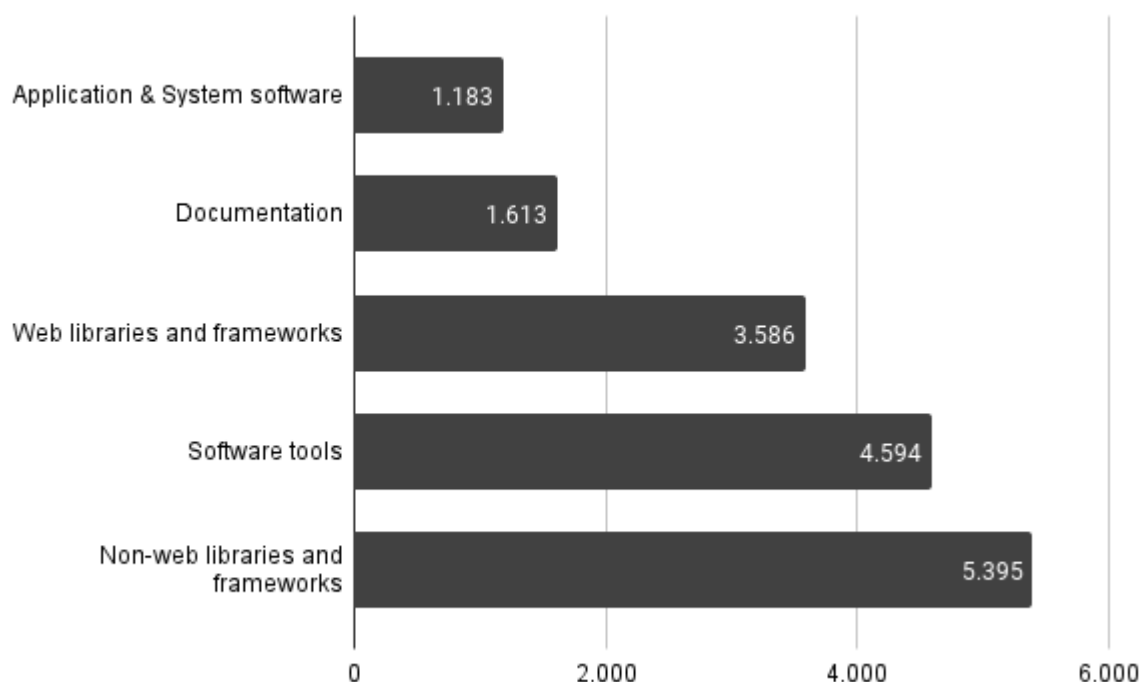
Diante desse cenário, podemos concluir que a adoção de projetos preexistentes não é uma prática comum entre as empresas que se envolvem com iniciativas OSS. Vale destacar que, durante nossa análise para identificar quais repositórios do conjunto de dados eram forks, descobrimos que um total de 845 repositórios não estão mais disponíveis. Isso afeta 388 empresas distintas, sendo a Microsoft a empresa com o maior número de repositórios deletados (81).

### 4.3. Quais são os domínios desses projetos?

Os resultados apresentados aqui referem-se aos 16.371 projetos que conseguimos classificar por domínio usando o classificador desenvolvido por Zanartu et al. . (Zanartu et al., 2022. No total, identificamos os seguintes números de projetos em cada categoria:

- 5.395 *Bibliotecas e Frameworks Não-Web*;
- 4.594 *Ferramentas de Software*;
- 3.586 *Bibliotecas e Frameworks Web*;
- 1.613 *Documentação*;
- 1.183 *Software de Aplicação e Sistema*.

A Fig. 2 mostra a distribuição relativa a cada domínio.



**Figura 2. Distribuição dos projetos por domínio.**

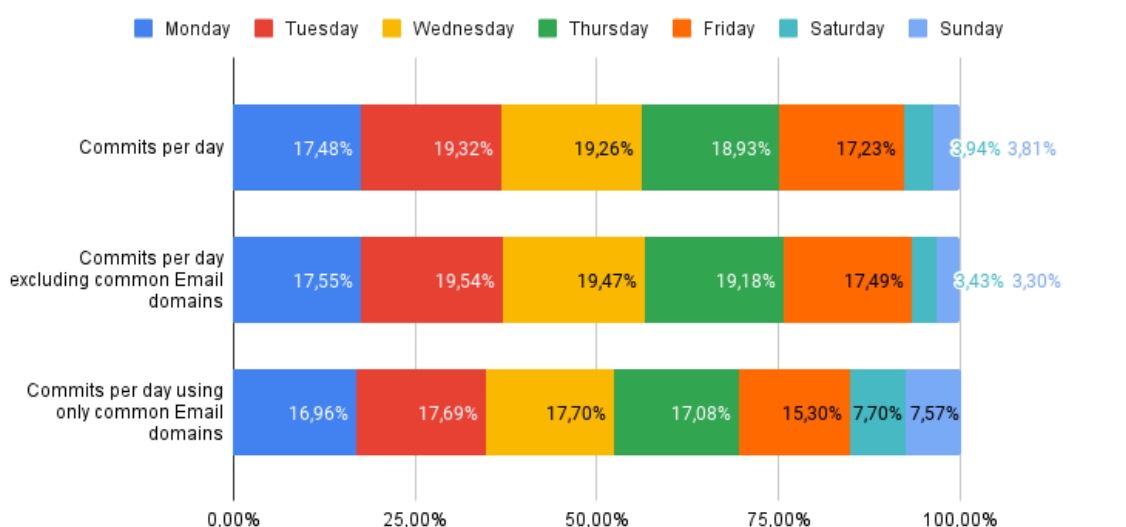
A primeira conclusão que podemos tirar é que bibliotecas, frameworks e ferramentas são, de longe, os tipos de projetos OSS nos quais as empresas mais se envolvem. Uma possível explicação é o movimento de empresas de software em direção à abertura de suas bibliotecas internas para compartilhá-las com a comunidade. Exemplos incluem o framework web React, da Meta, e a biblioteca de aprendizado de máquina TensorFlow, do Google.

Por outro lado, empresas raramente se envolvem com software de aplicação e sistemas, que representam apenas 7% do total de projetos OSS analisados. Uma possível explicação para isso é que as aplicações e sistemas das empresas são frequentemente seu diferencial competitivo, o que faz com que evitem disponibilizar seu código abertamente. Afinal, é improvável que o Google compartilhe o código-fonte de seu mecanismo de busca ou que a Uber disponibilize o código de seu aplicativo móvel.

#### 4.4. Como os commits de contribuidores de empresas e não empresas variam ao longo da semana?

A Fig. 3 mostra a distribuição dos commits nos repositórios, considerando três cenários:

- Commits de todos os domínios de e-mail;
- Commits excluindo domínios de e-mail comuns (como gmail.com, yahoo.com, etc.);
- Commits excluindo domínios comuns, ou seja, considerando apenas domínios empresariais (como microsoft.com, facebook.com, etc.).



**Figura 3. Quantidade de commits por dia da semana.**

Podemos observar que excluir domínios de e-mail comuns faz pouca diferença no número total de commits, o que já era esperado, pois a maioria dos commits nos repositórios apoiados por empresas são feitos com e-mails corporativos.

O conjunto de dados completo contém 7.851.155 commits, enquanto o subconjunto sem domínios de e-mail comuns contém 6.915.060 commits. Isso significa que os commits feitos com domínios comuns representam apenas cerca de 12% do total.

Além disso, vemos que a maior parte das contribuições para os repositórios apoiados por empresas ocorre durante os dias úteis, o que indica que esses repositórios são desenvolvidos principalmente por funcionários no horário de trabalho.



Por outro lado, quando consideramos apenas commits feitos com domínios de e-mail comuns, notamos um aumento no número de contribuições nos finais de semana. Isso sugere que esses commits são feitos por desenvolvedores voluntários, que contribuem para o projeto fora do horário comercial.

## 4.5. Quantos e quais projetos são apoiados por mais de uma empresa?

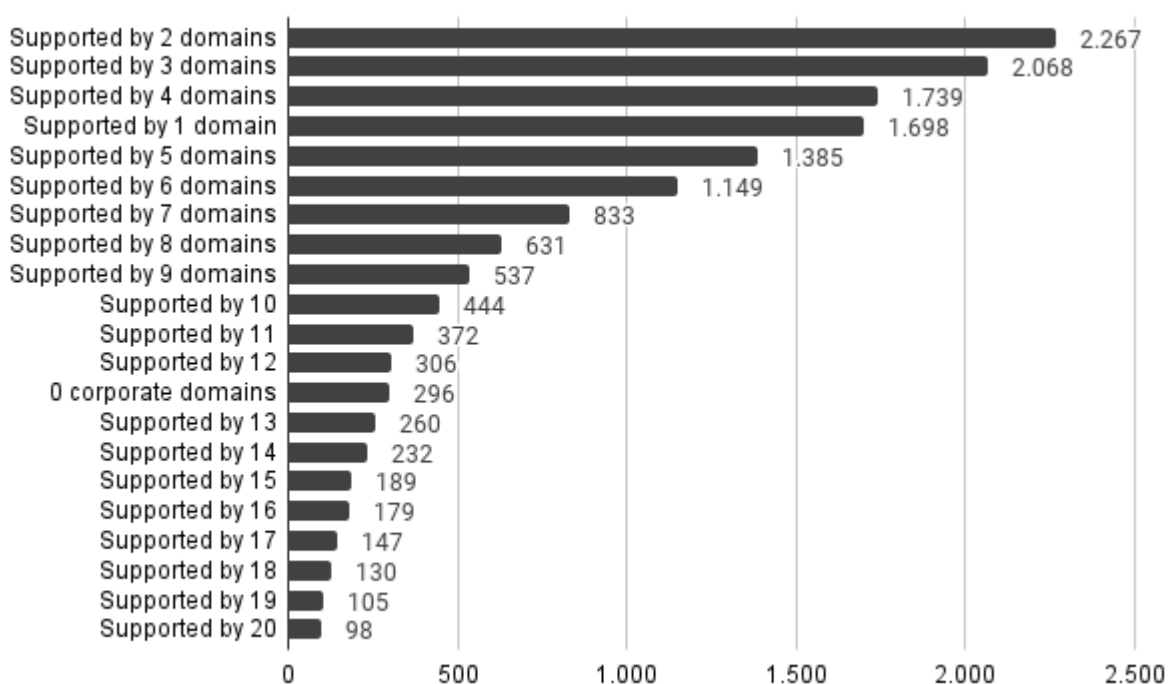
A maioria dos projetos é apoiada por um a seis domínios de e-mail diferentes. No total, 10.306 projetos distintos se enquadram nessa categoria, o que corresponde a aproximadamente 63% dos projetos.

O cenário mais comum é um projeto ser apoiado por dois domínios de e-mail diferentes, com 2.267 repositórios nessa categoria. Projetos apoiados por apenas um domínio vêm logo atrás, com 1.698 repositórios.

Vale ressaltar que domínios de e-mail únicos nem sempre indicam empresas únicas. Um projeto com dois domínios pode, na verdade, ser apoiado por uma única empresa (por exemplo, um funcionário contribuindo com o e-mail corporativo e outro com um e-mail pessoal).

A análise dos dados revela que, conforme nos aproximamos do quinto domínio de e-mail mais frequente em um repositório, a chance de encontrar domínios empresariais diminui significativamente.

A Fig. 4 mostra a distribuição do número de domínios de e-mail por projeto.



**Figura 4. Número de domínios de e-mail por projeto.**

## 5. Ameaças à Validade

Apesar dos esforços para garantir a precisão e confiabilidade dos resultados, este estudo possui algumas limitações. Inspirados em (Li et al., 2024 e (Nguyen-duc Daniela S. Cruzes, 2018), categorizamos essas ameaças em dois tipos principais: validade interna e validade de construção.

**Validade interna:** Como nosso conjunto de dados é baseado em informações publicamente disponíveis no GitHub, existe o risco de erros de classificação ou ausência de dados ao associar projetos a empresas. Embora tenhamos utilizado consultas Whois, a API do GitHub e verificações manuais para preencher os nomes das empresas, alguns repositórios podem estar incorretamente atribuídos ou omitidos.

Além disso, a propriedade corporativa não é estática. Fusões, aquisições e rebrandings podem levar a inconsistências na atribuição de empresas. Outro desafio é identificar a adoção de projetos preexistentes pelas empresas. Embora tenhamos analisado forks formais do GitHub e procurado forks informais nas descrições dos projetos, algumas empresas podem adotar código OSS de formas que não são facilmente detectáveis, levando a uma subestimação da taxa de adoção.

**Validade de construção:** Este estudo define o envolvimento das empresas com base nas contribuições feitas a partir de domínios de e-mail corporativos, abordagem utilizada com sucesso em outros estudos, como (Nguyen-duc Daniela S. Cruzes, 2018) e (Osborne et al., 2024).

Além disso, empregamos diversas técnicas para validar os dados, incluindo consultas Whois, a API do GitHub e análise manual. No entanto, algumas inconsistências menores ainda podem estar presentes.

## 6. Conclusão

Nossos resultados revelam que muitas empresas estão envolvidas com projetos OSS, o que demonstra uma mudança significativa no mercado de desenvolvimento de software, que antes era fortemente baseado em modelos de código fechado e proprietário.

Também observamos que, embora muitas empresas contribuam para projetos OSS, a maioria desses projetos é iniciada pela própria empresa. Além disso, as empresas tendem a se envolver com bibliotecas, frameworks e ferramentas, raramente abrindo o código de suas aplicações.

Por fim, nossos dados indicam que a maioria dos projetos OSS apoiados por empresas recebe suporte de no máximo seis domínios corporativos distintos, e que a colaboração entre diferentes empresas em um mesmo projeto ainda é limitada.

## Referências

BORGES, H.; HORA, A.; VALENTE, M. T. Understanding the factors that impact the popularity of github repositories. In: *2016 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. [S.l.: s.n.], 2016. p. 334–344. Disponível em:

<<https://ieeexplore.ieee.org/abstract/document/7816479>>. Acesso em: 12 jul. 2022. 3, 4

DIAS, L.; STEINMACHER, I.; PINTO, G. Who drives company-owned oss projects: internal or external members? *Journal of the Brazilian Computer Society*, Sociedade Brasileira de Computacao,

v. 24, n. 1, dez. 2018. ISSN 0104-6500. Disponível em:

<<https://link.springer.com/article/10.1186/s13173-018-0079-x>>. Acesso em: 8 fev. 2023. 2

FOUNDATION, T. L. *The Linux Foundation releases linux development report*. The Linux Foundation, 2015. Disponível em: <<https://www.linuxfoundation.org/press-release/the-linux-foundation-releases-linux-development-report/>>. Acesso em: 8 fev. 2023. 2

LI, X. et al. *Systematic Literature Review of Commercial Participation in Open Source Software*. 2024. Disponível em: <<https://arxiv.org/abs/2405.16880>>. Acesso em: 8 fev. 2023. 2, 8

NGUYEN-DUC DANIELA S. CRUZES, S. T. P. A. A. *Do software firms collaborate or compete? A model of coopetition in community initiated OSS projects*. 2018. Disponível em: <<https://arxiv.org/abs/1808.06489>>. Acesso em: 12 jul. 2022. 8, 9

OSBORNE, C. et al. *Characterising Open Source Co-opetition in Company-hosted Open Source Software Projects: The Cases of PyTorch, TensorFlow, and Transformers*. 2024. Disponível em: <<https://arxiv.org/abs/2410.18241>>. Acesso em: 18 nov. 2024. 2, 9

PINTO, G. et al. On the challenges of open-sourcing proprietary software projects. *Empirical Softw. Engg.*, Kluwer Academic Publishers, USA, v. 23, n. 6, p. 3221–3247, dec 2018. ISSN 1382-3256. Disponível em: <<https://doi.org/10.1007/s10664-018-9609-6>>. Acesso em: 18 nov. 2024. 1

PINTO, G.; STEINMACHER, I.; GEROSA, M. A. Common than you think: An in-depth study of casual contributors. In: *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*. [s.n.], 2016. v. 1, p. 112–123. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/7476635>>. Acesso em: 18 dez. 2024. 5

SPINELLIS, D. et al. A dataset of enterprise-driven open source software. In: *Proceedings of the 17th International Conference on Mining Software Repositories*. New York, NY, USA: Association for Computing Machinery, 2020. (MSR '20), p. 533–537. ISBN 9781450375177. Disponível em: <<https://doi.org/10.1145/3379597.3387495>>. Acesso em: 2 jan. 2024. 2, 4

ZANARTU, F. et al. *Automatically Categorising GitHub Repositories by Application Domain*. arXiv, 2022. Disponível em: <<https://arxiv.org/abs/2208.00269>>. Acesso em: 21 ago. 2024. 4, 6

ZHANG, Y. et al. How are paid and volunteer open source developers different? a study of the rust project. In: *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*. ACM, 2024. (ICSE '24), p. 1–13. Disponível em: <<http://dx.doi.org/10.1145/3597503.3639197>>. Acesso em: 3 Jun. 2024. 2

ZHANG, Y. et al. How do companies collaborate in open source ecosystems? an empirical study of openstack. In: *2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE)*. [s.n.], 2020. p. 1196–1208. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/3377811.3380376>>. Acesso em: 3 Jun. 2024. 2