



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Relatório Final - Laboratório 3
Laboratório de Experimentação de Software

Bruno Rocha Corrêa Urbano
Henrique Dani Franco Nezio
Pedro de Barros Alves

Belo Horizonte 2025

Introdução

Este trabalho tem como objetivo investigar como características específicas de uma PR podem influenciar o seu desfecho e o processo de revisão. Ao analisarmos PRs de repositórios populares, levantamos hipóteses informais com base em percepções empíricas sobre o comportamento da comunidade de desenvolvimento no GitHub.

As hipóteses informais levantadas pelo grupo foram as seguintes:

- **H1 (RQ01):** PRs maiores (com muitos arquivos modificados ou muitas linhas alteradas) têm maior chance de serem rejeitadas, pois revisores podem ter mais dificuldade em revisar mudanças extensas com segurança.
- **H2 (RQ02):** PRs que demoram muito tempo para serem analisadas têm maior chance de serem rejeitadas, talvez por falta de engajamento, conflitos com a base principal ou por dúvidas não resolvidas no processo.
- **H3 (RQ03):** PRs com descrições mais completas e detalhadas são mais facilmente compreendidas pelos revisores, o que aumenta a probabilidade de aprovação (merge).
- **H4 (RQ04):** PRs com mais interações (comentários, participantes) indicam um processo mais colaborativo e, por isso, tendem a ser mais bem avaliadas e aprovadas ao final.
- **H5 (RQ05):** PRs maiores requerem mais revisões, pois envolvem mais pontos de atenção e discussões detalhadas por parte dos revisores.
- **H6 (RQ06):** PRs que levam mais tempo para serem fechadas ou mescladas geralmente passam por mais revisões, especialmente se houver muitas iterações de comentários e sugestões.
- **H7 (RQ07):** PRs com descrições mais completas e claras tendem a demandar menos revisões, pois reduzem mal-entendidos e facilitam a compreensão das mudanças propostas.

- **H8 (RQ08):** PRs com muitas interações geralmente passam por mais revisões, pois envolvem ciclos de feedback entre autores e revisores para alinhar mudanças, resolver dúvidas ou solicitar melhorias.

Metodologia

A metodologia adotada para este projeto consistiu nas seguintes etapas:

1. Seleção dos Repositórios:

Inicialmente, foi realizada a coleta dos 300 repositórios mais populares do GitHub, utilizando a API GraphQL. A seleção foi feita a partir do critério de popularidade (número de estrelas) e os dados básicos dos repositórios foram armazenados, incluindo nome, dono, quantidade de estrelas, data de criação, data de atualização e URL.

2. Filtro dos Repositórios:

Após a coleta inicial, foram selecionados apenas os repositórios que possuíam pelo menos 100 Pull Requests (soma de PRs com status Merged e Closed), assegurando que a análise focasse em projetos com atividade significativa de colaboração.

3. Coleta de Pull Requests:

Para cada repositório filtrado, foram coletadas as Pull Requests fechadas utilizando a GitHub REST API. Aplicou-se o seguinte conjunto de critérios de inclusão:

- PRs com status Merged ou Closed;
- PRs que permaneceram abertas por mais de uma hora (tempo entre a criação e o fechamento/merge);
- PRs que passaram por pelo menos uma revisão formal (excluindo PRs sem revisões).

4. Extração de Métricas:

De cada PR selecionada, foram extraídas as seguintes informações:

- Tempo de Análise: Duração, em horas, entre a criação e o fechamento/merge da PR.
- Tamanho: Número de arquivos modificados, linhas adicionadas e linhas removidas.
- Descrição: Comprimento (número de caracteres) do corpo (body) da PR.
- Interações: Número de comentários, número de comentários de revisão, e número de participantes únicos envolvidos na PR.
- Outros Dados: Estado final da PR (Merged ou Closed), autor da PR, e informações básicas do título.

5. Construção do Dataset Final:

Todas as métricas extraídas foram organizadas em um único conjunto de dados tabular (arquivo .csv), que serviu como base para responder às questões de pesquisa propostas.

6. Análise Estatística:

A análise dos dados foi realizada considerando os valores medianos das métricas para todos os PRs coletados. Posteriormente, foram elaborados gráficos, correlações e interpretações visando responder às questões de pesquisa, buscando validar ou refutar as hipóteses informais levantadas inicialmente.