

# Caracterizando a Atividade de Code Review no GitHub

Matheus Machado de Oliveira Andrade<sup>1</sup>, Rafael Parreira Chequer<sup>2</sup>, Samara Martins Ferreira<sup>3</sup>, Victor Alexandre Peters Fonseca<sup>4</sup>

Instituto de Ciências Exatas e Informática  
Pontifícia Universidade de Minas Gerais (PUC Minas)  
Belo Horizonte – MG – Brasil

{500391, 1275759, 1328767, 1314281}@sga.pucminas.br

**Resumo.** *Este laboratório, fará uma análise da atividade de **code review** em repositórios populares do **GitHub**. O foco será identificar as variáveis que influenciam na decisão de **merge** de um **Pull Request (PR)**. A análise será feita sob a perspectiva dos desenvolvedores que submetem código aos repositórios selecionados. Isso envolve a inspeção do código produzido antes de sua integração à base principal, garantindo a qualidade do código e evitando a inclusão de defeitos. Além disso, será considerado o papel das ferramentas de verificação estática que avaliam requisitos de estilo de programação ou padrões definidos pela organização.*

## 1. Introdução

A prática de code review tornou-se uma constante nos processos de desenvolvimento ágeis. Em linhas gerais, ela consiste na interação entre desenvolvedores e revisores visando inspecionar o código produzido antes de integrá-lo à base principal. Assim, garante-se a qualidade do código integrado, evitando-se também a inclusão de defeitos. No contexto de sistemas open source, mais especificamente dos desenvolvidos através do GitHub, as atividades de code review acontecem a partir da avaliação de contribuições submetidas por meio de Pull Requests (PR). Ou seja, para que se integre um código na branch principal, é necessário que seja realizada uma solicitação de pull, que será avaliada e discutida por um colaborador do projeto. Ao final, a solicitação de *merge* pode ser aprovada ou rejeitada pelo revisor. Em muitos casos, ferramentas de verificação estática realizam uma primeira análise, avaliando requisitos de estilo de programação ou padrões definidos pela organização.

Neste contexto, o objetivo deste laboratório é analisar a atividade de code review desenvolvida em repositórios populares do GitHub, identificando variáveis que influenciam no merge de um PR, sob a perspectiva de desenvolvedores que submetem código aos repositórios selecionados.

### 1.1. Questões de Pesquisa e Hipóteses Informais

#### A. Feedback Final das Revisões (Status do PR):

**RQ 01 - Qual a relação entre o tamanho dos PRs e o feedback final das revisões?**

O tamanho dos Pull Requests pode influenciar o feedback final das revisões. PRs muito grandes podem ser mais difíceis de revisar e, portanto, podem ter mais chances de serem rejeitados.

**RQ 02 - Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?**

O tempo de análise dos PRs também pode afetar o feedback final. PRs que levam muito tempo para serem analisados podem indicar problemas complexos, o que pode levar a um feedback negativo.

**RQ 03 - Qual a relação entre a descrição dos PRs e o feedback final das revisões?**

A descrição dos PRs é crucial para o feedback final. Descrições claras e detalhadas podem facilitar o processo de revisão e aumentar as chances de aprovação do PR.

**RQ 04 - Qual a relação entre as interações nos PRs e o feedback final das revisões?**

As interações nos PRs, como comentários e discussões, podem influenciar o feedback final. PRs com muitas interações podem indicar um alto nível de colaboração, o que pode levar a um feedback mais positivo.

**B. Número de Revisões:**

**RQ 05 - Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?**

O tamanho dos Pull Requests pode influenciar o número de revisões realizadas. Pull Requests maiores podem exigir mais revisões para garantir a qualidade do código.

**RQ 06 - Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?**

O tempo de análise dos PRs pode afetar o número de revisões. Pull Requests que levam mais tempo para serem analisados podem exigir mais revisões.

**RQ 07 - Qual a relação entre a descrição dos PRs e o número de revisões realizadas?**

A descrição dos PRs pode influenciar o número de revisões. PRs com descrições claras e detalhadas podem facilitar o processo de revisão e, portanto, exigir menos revisões.

**RQ 08 - Qual a relação entre as interações nos PRs e o número de revisões realizadas?**

As interações nos PRs podem afetar o número de revisões. PRs com muitas interações podem indicar um alto nível de colaboração e, portanto, podem exigir mais revisões para considerar todas as opiniões e sugestões.

## **2. Metodologia**

### **2.1. Definição do Problema**

O problema em questão envolve a análise da atividade de code review em repositórios populares do GitHub, com o objetivo de identificar as variáveis que influenciam na decisão de mesclar um Pull Request (PR). Isso implica entender como os desenvolvedores submetem e revisam código, e quais fatores levam à aprovação ou rejeição de uma contribuição.

### **2.1. Formulação de Hipóteses**

- **Hipótese 1:** O número de estrelas de um repositório está correlacionado com a quantidade e a qualidade dos Pull Requests submetidos.
- **Hipótese 2:** A quantidade de revisões realizadas em um Pull Request está relacionada à sua probabilidade de ser mesclado.
- **Hipótese 3:** A presença de revisões automatizadas, como análise estática de código, influencia na probabilidade de um Pull Request ser mesclado.

### **2.2. Coleta de Dados**

A coleta de dados é realizada através da API do GitHub, utilizando consultas GraphQL para obter informações sobre repositórios, Pull Requests e suas respectivas características. São coletados dados como nome do repositório, proprietário, número de estrelas, número total de Pull Requests, entre outros.

### **2.3. Processamento e Análise de Dados**

Os dados coletados são processados e analisados para extrair informações relevantes para responder às hipóteses formuladas. Isso inclui a criação de um conjunto de dados estruturado, onde são calculadas métricas como o número de estrelas, o total de Pull Requests, o número de revisões realizadas, e a presença de revisões automatizadas. Esses dados são então analisados estatisticamente para identificar padrões e correlações entre as variáveis, ajudando a compreender quais fatores influenciam na decisão de mesclar um PR.

## **2.4. Interpretação dos Resultados**

Inserir aqui a interpretação dos resultados

## **3. Resultados**

Inserir aqui os resultados

## **4. Discussão**

Discutir os resultados

## **5. Conclusões**

Conclusão do trabalho