

# **Caracterizando a Atividade de Code Review em Repositórios Populares do GitHub**

Luiz Filipe Nery Costa  
*(PUC MINAS)*

Wilken Henrique Moreira  
*PUC MINAS*

16 de outubro de 2025

# Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Metodologia</b>	<b>3</b>
2.1	Criação do Dataset . . . . .	3
2.2	Questões de Pesquisa (RQs) . . . . .	3
2.3	Métricas de Pull Request . . . . .	4
2.4	Desafios Enfrentados . . . . .	4
<b>3</b>	<b>Resultados</b>	<b>4</b>
3.1	RQ 01: Qual a relação entre o tamanho dos PRs e o seu feedback final? . . . . .	4
3.2	RQ 02: Qual a relação entre o tempo de análise dos PRs e o seu feedback final? . . . . .	5
3.3	RQ 03: Qual a relação entre a descrição dos PRs e o feedback final das revisões? . . . . .	6
3.4	RQ 04: Qual a relação entre as interações nos PRs e o feedback final das revisões? . . . . .	7
3.5	RQ 05: Qual a relação entre o tamanho dos PRs e o número de revisões realizadas? . . . . .	8
3.6	RQ 06: Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas? . . . . .	9
3.7	RQ 07: Qual a relação entre a descrição dos PRs e o número de revisões realizadas? . . . . .	10
3.8	RQ 08: Qual a relação entre as interações nos PRs e o número de revisões realizadas? . . . . .	11
<b>4</b>	<b>Conclusão</b>	<b>12</b>

# 1 Introdução

A prática de *code review* é um pilar no desenvolvimento de software moderno, especialmente em projetos open source na plataforma GitHub. Este estudo analisa um dataset de Pull Requests (PRs) dos 200 repositórios mais populares para identificar os fatores que influenciam o seu desfecho.

Em linhas gerais, o processo de code review 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 no GitHub, as atividades de *code review* acontecem a partir da avaliação de contribuições submetidas por meio de Pull Requests (PRs). 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.

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.

## 2 Metodologia

### 2.1 Criação do Dataset

O dataset utilizado foi composto pelos **últimos 300 Pull Requests** de repositórios que atendem aos seguintes critérios:

- **Populares:** PRs submetidos aos 200 repositórios mais populares do GitHub.
- **Ativos:** Repositórios com pelo menos 100 PRs (MERGED + CLOSED).

A fim de analisar PRs que passaram por uma revisão humana, foram selecionados apenas aqueles com as seguintes características:

- Com status MERGED ou CLOSED.
- Que possuam pelo menos uma revisão.
- Cuja revisão levou pelo menos uma hora.

### 2.2 Questões de Pesquisa (RQs)

O estudo buscou responder às seguintes questões:

- **RQ 01:** Qual a relação entre o **tamanho** dos PRs e o seu feedback final?
- **RQ 02:** Qual a relação entre o **tempo de análise** dos PRs e o seu feedback final?
- **RQ 03:** Qual a relação entre a **descrição** dos PRs e o seu feedback final?
- **RQ 04:** Qual a relação entre as **interações** nos PRs e o seu feedback final?
- **RQs 05:** Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?
- **RQs 06:** Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?
- **RQs 07:** Qual a relação entre a descrição dos PRs e o número de revisões realizadas?

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

## 2.3 Métricas de Pull Request

Para avaliar objetivamente o processo de revisão, este estudo utilizou um conjunto de métricas que quantificam dimensões distintas de um Pull Request:

- **Tamanho:** Mede a magnitude da contribuição. É quantificado pelo número de arquivos modificados e pelo total de linhas de código adicionadas e removidas.
- **Tempo de Análise:** Representa o tempo que um PR permaneceu em discussão. É definido como o intervalo entre a data de criação do PR e a data de sua última atividade (seja o merge ou o fechamento).
- **Descrição:** Avalia o esforço do contribuidor em contextualizar sua mudança. É mensurado pelo número de caracteres presentes no corpo da descrição do PR.
- **Interações:** Captura o nível de engajamento e discussão gerado pelo PR. É medido pelo número de participantes envolvidos na discussão e pelo número total de comentários.

## 2.4 Desafios Enfrentados

Durante a execução, alguns desafios foram encontrados:

- **Tempo de Execução das Buscas:** A coleta de dados através da API do GitHub para 200 repositórios representou um desafio computacional.
- **Repositórios com Números Massivos de PRs:** Foi adotada uma estratégia de amostragem, limitando a coleta aos 300 PRs mais recentes de cada projeto.

# 3 Resultados

## 3.1 RQ 01: Qual a relação entre o tamanho dos PRs e o seu feedback final?

A análise dos dados para a RQ01 revela uma forte e consistente relação negativa entre o tamanho de um Pull Request e sua probabilidade de aceitação. Os PRs aceitos (MERGED) são significativamente menores (mediana de 4.0 arquivos e 84.1 linhas alteradas) do que os rejeitados (CLOSED), que apresentam medianas de 6.0 arquivos e 186.1 linhas.

A preferência por contribuições menores reflete a redução da carga cognitiva para os revisores e o menor risco associado a mudanças mais focadas. PRs extensos são mais difíceis de validar e tendem a gerar discussões mais longas, o que aumenta a probabilidade de rejeição. A grande maioria das contribuições bem-sucedidas são, portanto, pequenas e focadas em um único objetivo.

A preferência por PRs menores não é uma mera coincidência, mas um reflexo direto dos desafios práticos e cognitivos inerentes ao processo de code review. As principais razões que explicam este fenômeno são:

- **Carga Cognitiva do Revisor**
- **Risco Associado à Mudança**

- **Clareza de Propósito e Facilidade de Discussão**

RQ01: Relação entre o Tamanho do PR e o Status Final (Com Zoom)

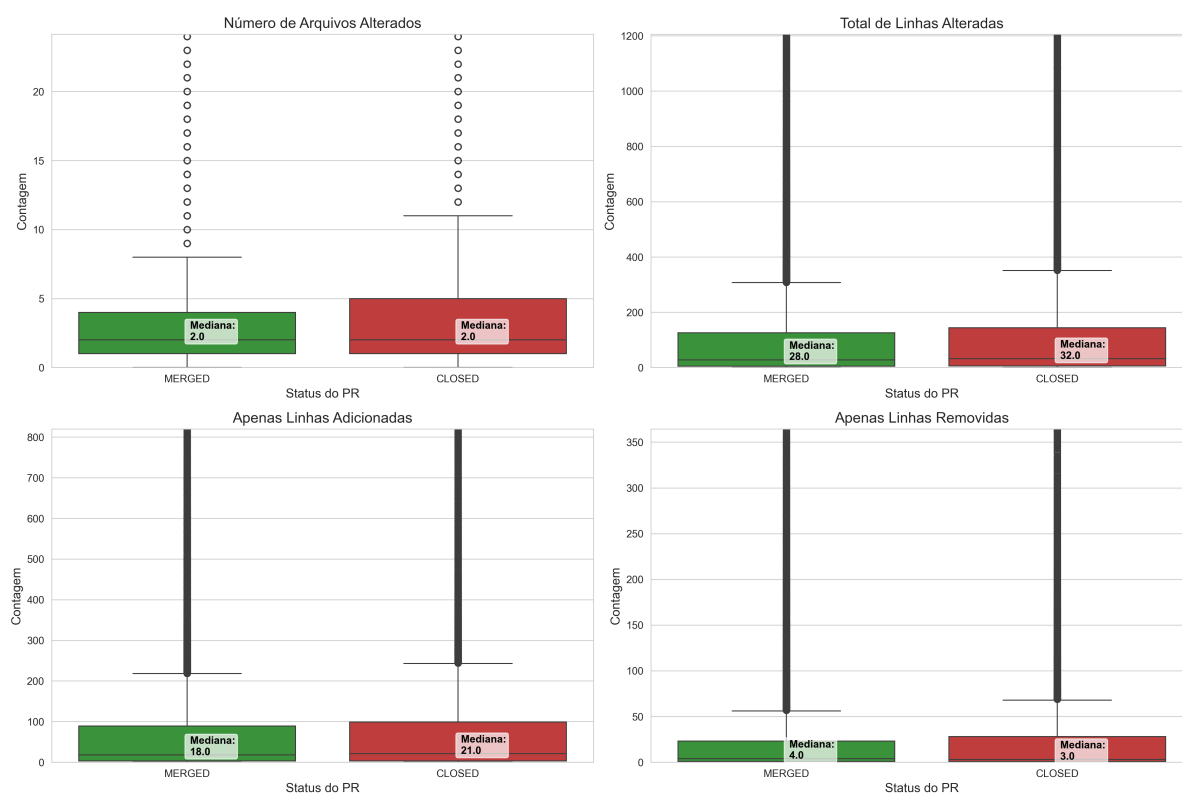


Figura 1: Relação entre as métricas de tamanho e o status final do PR.

### 3.2 RQ 02: Qual a relação entre o tempo de análise dos PRs e o seu feedback final?

A segunda questão de pesquisa investiga a influência da agilidade do processo de revisão no desfecho de um Pull Request. A análise dos dados demonstra que o tempo de análise é um indicador crítico, com uma forte relação negativa entre a duração do processo e a probabilidade de aceitação da contribuição. Em suma, PRs que são resolvidos rapidamente têm uma chance muito maior de serem integrados.

A diferença no comportamento temporal entre os PRs aceitos e rejeitados é drástica: Tempo de Análise Mediano: Um PR aceito (MERGED) leva, na mediana, 33.5 horas para ser concluído. Em contraste, a mediana para um PR rejeitado (CLOSED) é de 95.8 horas – quase o triplo do tempo.

- **Momentum e Engajamento:** Um PR que recebe feedback rápido mantém o "momentum", mantendo o autor engajado e motivado a responder. Processos demorados podem levar ao desengajamento e até ao abandono da contribuição.
- **Sinal de Clareza e Consenso:** A rapidez na decisão frequentemente sinaliza que a proposta era clara, bem-feita e alinhada com os objetivos do projeto desde o início, indicando que houve pouco debate ou controvérsia para chegar a um consenso.
- **Risco de Obsolescência (Code Rot) :** Em projetos ativos, a base de código evolui constantemente.

Um PR que permanece aberto por muito tempo corre um risco crescente de entrar em conflito com novas mudanças, tornando sua integração mais complexa e aumentando a chance de ser descartado.

- **Interconexão com o Tamanho (RQ01):** Este resultado está diretamente ligado à nossa conclusão anterior. PRs menores são inerentemente mais rápidos de revisar. A agilidade, portanto, não é apenas uma causa do sucesso, mas também uma consequência do fato de a contribuição ser pequena e focada.

RQ02: Relação entre o Tempo de Análise e o Status Final (Com Zoom)

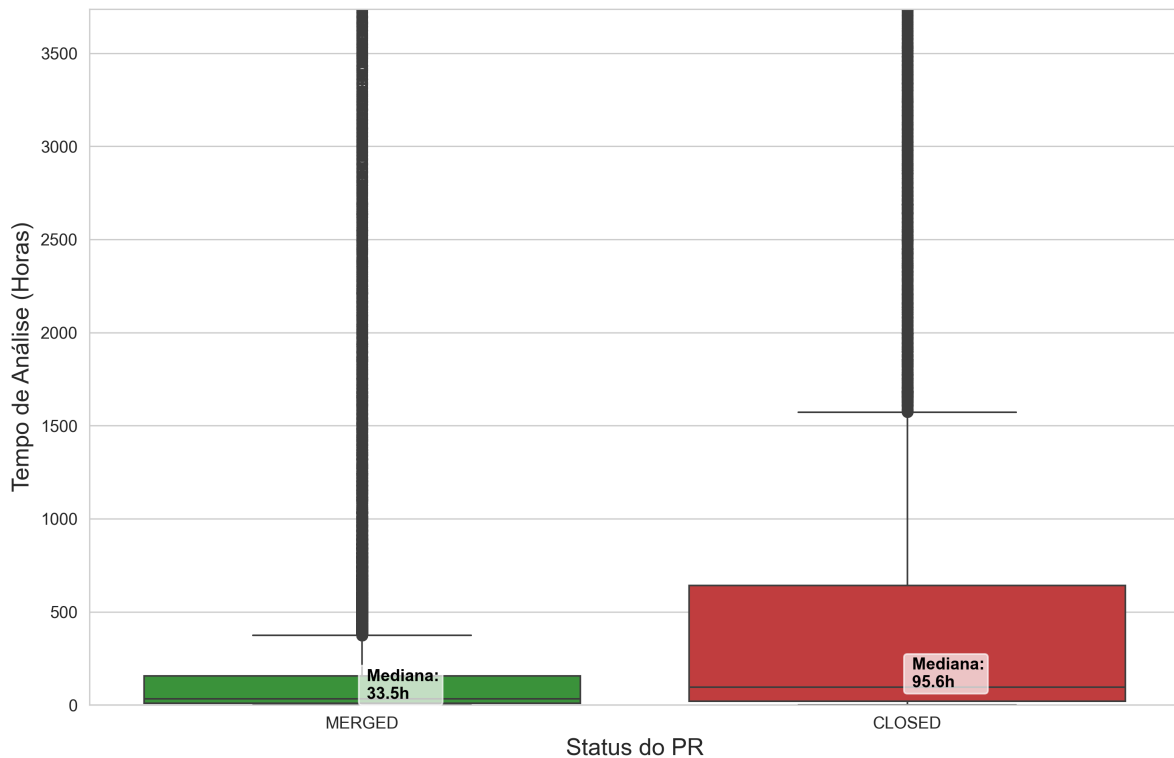


Figura 2: Relação entre as métricas de tamanho e o status final do PR.

### 3.3 RQ 03: Qual a relação entre a descrição dos PRs e o feedback final das revisões?

A análise da RQ03 demonstra que a comunicação escrita inicial é um fator decisivo no desfecho de um Pull Request. Os dados mostram uma forte relação positiva entre o detalhe da descrição e a probabilidade de aceitação.

PRs aceitos (MERGED) possuem descrições com mais do que o dobro do tamanho (mediana de 333 caracteres) em comparação com os rejeitados (CLOSED), que apresentam uma mediana de apenas 152 caracteres.

Isso indica que uma descrição detalhada é um forte indicador de sucesso. Ela funciona como um "manual" para o revisor, reduzindo sua carga cognitiva, contextualizando a mudança e sinalizando o empenho do contribuidor. Uma descrição vaga ou ausente, por outro lado, é um sinal de alerta (red flag) que aumenta a fricção no processo e a probabilidade de rejeição, pois força os revisores a um trabalho de adivinhação para entender o propósito da mudança.

- **Redução da Carga Cognitiva do Revisor:** Uma boa descrição contextualiza o problema (por que a mudança é necessária), explica a solução implementada (como ela funciona) e, muitas vezes, sugere como testá-la. Isso economiza um tempo valioso do revisor, que não precisa "adivinhar" a intenção do autor ou fazer engenharia reversa da lógica do código.
- **Sinal de Profissionalismo e Empenho:** Um contribuidor que investe tempo para escrever uma descrição clara demonstra profissionalismo e respeito pelo tempo dos revisores. Isso cria uma impressão positiva e sinaliza que o autor provavelmente também foi cuidadoso ao escrever o código.
- **Antecipação de Dúvidas:** Descrições detalhadas podem antecipar e responder a perguntas que os revisores fariam de qualquer maneira. Isso previne longas threads de comentários e acelera o ciclo de feedback, contribuindo para um tempo de análise menor (como vimos na RQ02).
- **Documentação Histórica:** A descrição de um PR aceito torna-se parte da história do projeto. Quando um desenvolvedor, meses ou anos depois, usa comandos como git blame para entender por que uma linha de código existe, a descrição do PR associado fornece um contexto crucial. PRs com descrições pobres deixam uma "dívida" de documentação no projeto.

Em resumo, para a RQ03, os dados confirmam que a comunicação escrita é um pilar do processo de code review. Uma descrição vaga ou ausente é um forte sinal de alerta (red flag) que aumenta a fricção e a probabilidade de rejeição. Investir tempo na elaboração de uma descrição clara e completa é uma das ações de maior retorno para aumentar as chances de um PR ser aceito.

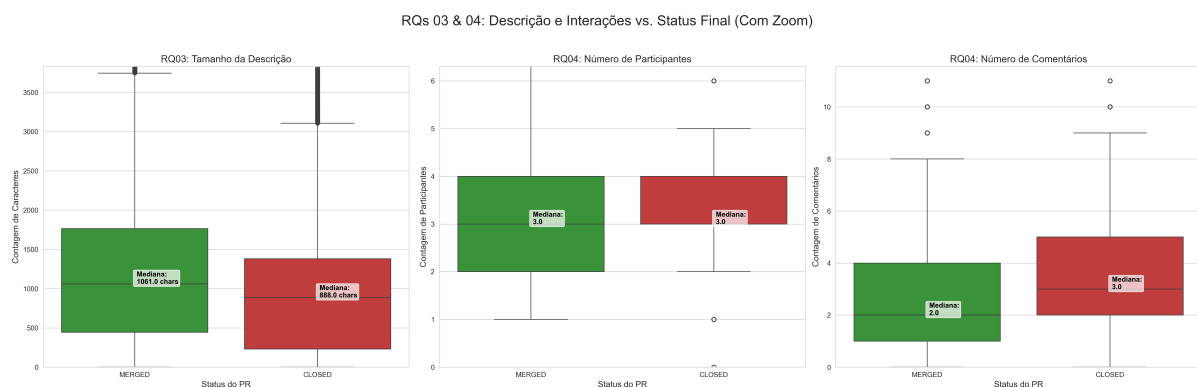


Figura 3: Relação entre as métricas de tamanho e o status final do PR.

### 3.4 RQ 04: Qual a relação entre as interações nos PRs e o feedback final das revisões?

A análise dos dados revela que o volume de discussão em um Pull Request é um forte indicador que existe uma relação negativa entre o número de interações (tanto participantes quanto comentários) e a probabilidade de um PR ser aceito.

Embora a colaboração seja central para o open source, um volume excessivo de interações num PR específico frequentemente não é um sinal de colaboração saudável, mas sim um sintoma de problemas subjacentes.

Uma observação pertinente é que ao olhar para o número de comentários de um PR aceite (MERGED) tem uma mediana de 2.0 participantes (tipicamente o autor e um revisor). Em contraste, um PR rejeitado

(CLOSED) tem uma mediana de 3.0 participantes, ou seja, houve a necessidade de envolver mais pessoas na discussão de um PR é um sinal de que a contribuição pode ser controversa, complexa ou impactar múltiplas áreas do projeto, aumentando a dificuldade de se chegar a um consenso.

A mediana de comentários para um PR aceite (MERGED) é de 3.0. Para um PR rejeitado (CLOSED), a mediana salta para 5.0, sendo assim, um PR bem-sucedido gera, tipicamente, poucas discussões. O fluxo ideal parece ser: o autor submete, o revisor faz alguns comentários pontuais, o autor ajusta e o PR é aceite. Um número elevado de comentários indica um ciclo de revisão mais longo e com mais atrito, onde podem estar a ser discutidos problemas fundamentais de design, bugs encontrados ou desalinhamento com as diretrizes do projeto.

- **Eficiência vs. Controvérsia:** Um baixo número de interações geralmente significa que o PR era claro, bem implementado e consensual. A revisão é rápida e o processo flui sem obstáculos. Um alto número de interações, por outro lado, é frequentemente um sintoma de problemas: a solução pode ser controversa, o código pode ter bugs, a abordagem pode não ser a ideal ou a descrição pode ser vaga (ligando-se à RQ03).
- **Sinal de Alerta:** Um volume crescente de comentários e a entrada de novos participantes numa discussão de PR devem ser vistos como um sinal de alerta pelos mantenedores do projeto. Indica que o PR está a consumir uma quantidade desproporcional de tempo e atenção da equipa, o que pode justificar uma intervenção para decidir se vale a pena continuar a investir esforço nele ou se seria mais produtivo fechá-lo e pedir ao autor uma nova abordagem.
- **Interconexão com Outras RQs:** Esta questão está intrinsecamente ligada às outras. PRs maiores (RQ01) naturalmente atraem mais comentários e participantes. Discussões mais longas, por sua vez, aumentam o tempo de análise (RQ02). Vemos aqui um ciclo vicioso: um PR grande leva a mais discussões, o que leva a um processo mais longo, o que, por fim, aumenta a probabilidade de rejeição.

O volume de interações funciona como um "termómetro" para a saúde de um PR. Enquanto alguma discussão é necessária e saudável, um excesso de debate é um forte indicador de que a contribuição está em risco de ser rejeitada

### 3.5 RQ 05: Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?

A análise da RQ05, que investiga a relação entre o tamanho de um PR e o esforço de revisão, revela uma correlação positiva moderada. Os dados indicam que, de forma geral, quanto maior a contribuição, mais comentários e discussões ela tende a gerar.

A força da correlação entre o tamanho do PR (medido pelo número de arquivos) e o volume de comentários é de 0.40.

- **Aumento da Superfície de Análise:** Cada arquivo e cada linha de código adicionada ou modificada é um ponto potencial para a introdução de bugs, problemas de performance ou desalinhamento com a arquitetura do projeto. PRs maiores simplesmente têm uma "superfície de análise" maior, o que naturalmente leva a um maior número de observações e perguntas por parte dos revisores.



- **Necessidade de Múltiplos Especialistas:** Um PR extenso frequentemente toca em diferentes áreas do sistema (ex: front-end, back-end, base de dados). Isso exige a opinião de múltiplos revisores, cada um especialista em uma área. Como vimos na RQ04, mais participantes levam a mais comentários. Portanto, o tamanho atua como um gatilho para aumentar o número de pessoas envolvidas.
- **Complexidade e Incerteza:** A complexidade de uma mudança tende a crescer exponencialmente com o seu tamanho. Para os revisores, entender as implicações de uma grande mudança é mais difícil e gera mais incerteza. Essa incerteza se manifesta em forma de perguntas, pedidos de clarificação e sugestões de abordagens alternativas, inflando o número de comentários.

Um PR grande não apenas tem uma chance menor de ser aceite (como vimos na RQ01), mas também consome uma quantidade desproporcional do tempo e da atenção da equipa de revisão. Este resultado reforça, mais uma vez, a recomendação de quebrar grandes funcionalidades em contribuições menores para tornar o processo de revisão mais eficiente e gerenciável.

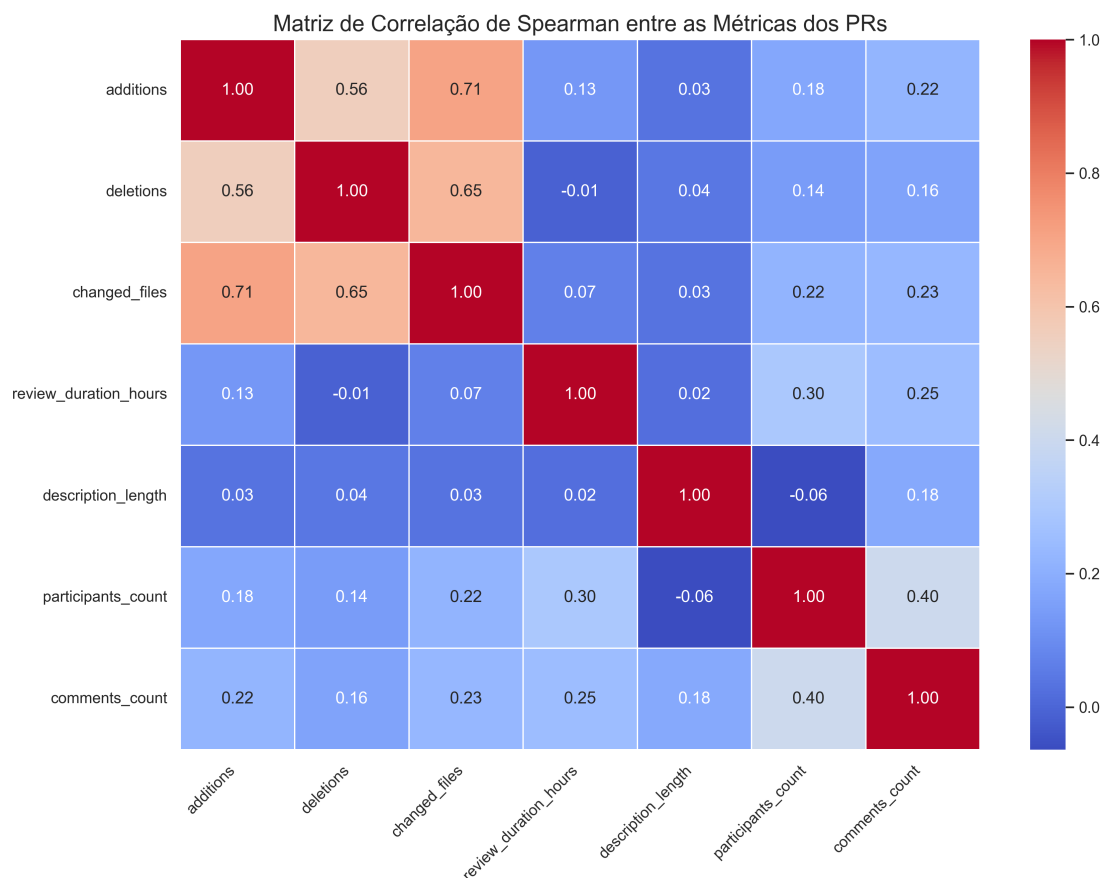


Figura 4: Relação entre as correlação das métricas.

### 3.6 RQ 06: Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?

A análise dos dados revela uma correlação positiva moderada entre o tempo que um Pull Request permanece aberto e o volume de discussões que ele gera. Em outras palavras, PRs com um ciclo de vida

mais longo tendem a acumular um número maior de comentários.

O gráfico mostra que a força da correlação de Spearman entre o Tempo de Análise (review duration hours) e o número de comentários é de 0.40, esse valor, idêntico ao da correlação com o tamanho do PR, indica uma relação moderada e significativa. Fica claro que a duração do processo de revisão está diretamente associada à intensidade do debate.

- **Mais Comentários Levam a Mais Tempo:** Este é o fluxo mais intuitivo. Um PR que gera muitos comentários e discussões naturalmente levará mais tempo para ser resolvido. Cada comentário representa um ponto a ser discutido, uma mudança a ser implementada ou uma dúvida a ser esclarecida. Cada ciclo de "revisor comenta - autor ajusta - novo review" adiciona um tempo considerável ao ciclo de vida total do PR.
- **Mais Tempo Leva a Mais Comentários:** O fluxo inverso também é verdadeiro. Um PR que permanece aberto por um longo período—seja por falta de atenção dos revisores, por ser de baixa prioridade ou por estar bloqueado por outras dependências—tem uma janela de oportunidade maior para acumular discussões. Novos revisores podem encontrá-lo, ou mudanças na base de código principal podem levantar novas questões sobre a relevância e a implementação do PR, gerando novos comentários.

As PRs grandes (RQ01) geram mais discussões (RQ05), e essas discussões, por sua vez, estendem o tempo de análise (esta RQ06). Todos esses fatores, em conjunto, aumentam a probabilidade de rejeição do PR (como vimos nas RQs da primeira dimensão). Um tempo de análise longo não é apenas uma métrica isolada; é frequentemente um sintoma de um PR complexo, controverso ou que não está a receber a atenção necessária para progredir.

### 3.7 RQ 07: Qual a relação entre a descrição dos PRs e o número de revisões realizadas?

A análise revela que a relação entre o tamanho da descrição de um PR e o volume de discussões que ele gera é a mais fraca entre todos os fatores analisados. Os dados mostram uma correlação positiva muito fraca, indicando que o número de caracteres na descrição tem pouca influência na quantidade de comentários.

O gráfico de correlações mostra que a força da correlação de Spearman entre o Tamanho da Descrição e o número de comentários é de apenas **0.19**.

- **Qualidade vs. Quantidade:** Este resultado sugere que, embora uma descrição detalhada seja crucial para a aceitação de um PR (como visto na RQ03), ela não necessariamente reduz o volume de discussões técnicas. Uma mudança complexa, mesmo que bem descrita, ainda irá gerar um debate aprofundado sobre sua implementação.
- **Mudança no Foco da Discussão:** Uma boa descrição pode mudar o foco dos comentários. Em vez de perguntas básicas como "O que este PR faz?", a discussão pode evoluir para questões mais técnicas e aprofundadas, como "Você considerou este caso de teste específico?". A quantidade de comentários pode permanecer a mesma, mas a qualidade da discussão melhora.

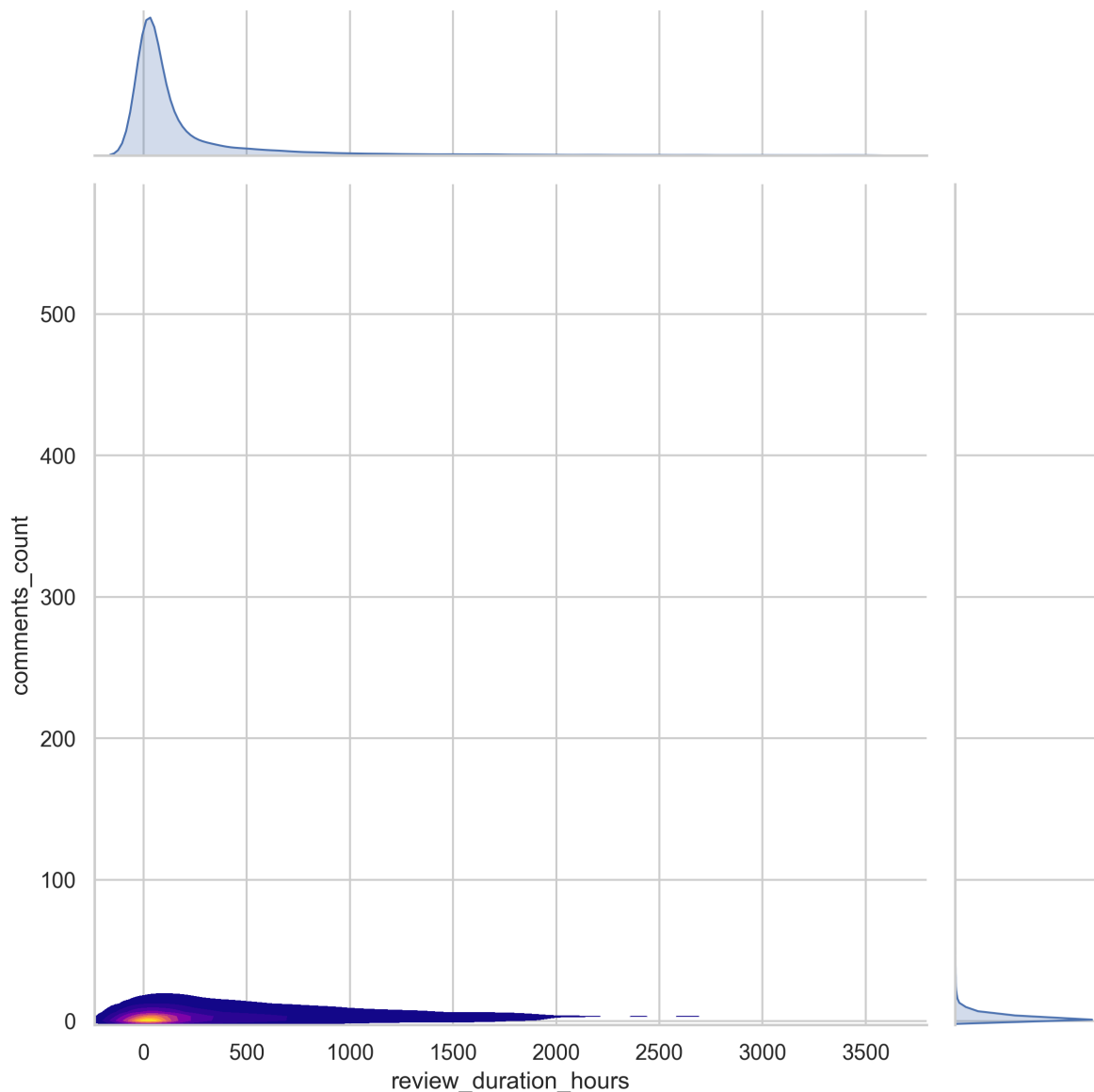


Figura 5: Relação entre as correlação das métricas.

### 3.8 RQ 08: Qual a relação entre as interações nos PRs e o número de revisões realizadas?

A análise da RQ08 mostra a relação mais forte e intuitiva de toda a segunda dimensão da pesquisa. Existe uma **\*\*correlação positiva extremamente forte\*\*** entre o número de participantes em um PR e o número total de comentários. O número de pessoas envolvidas é, de longe, o melhor preditor para o volume de discussões.

O gráfico de correlações quantifica esta relação de forma clara, mostrando um coeficiente de Spearman de **0.75** entre Participantes e Comentários.

- **Complexidade Social da Revisão:** Este resultado destaca a "complexidade social" de um processo de revisão. Cada novo participante é uma nova fonte de feedback, perguntas e sugestões, o que naturalmente infla o número de comentários.
- **Sinal de Interdependência:** Um PR que exige a opinião de múltiplos stakeholders (ex: especialistas

de front-end, back-end, segurança e design) terá, inerentemente, um volume de discussão muito maior. O número de participantes, portanto, funciona como um excelente proxy para a complexidade e a interdependência da mudança proposta.

- **Ligação com o Risco de Rejeição:** Este achado reforça o que foi visto na RQ04. Um alto número de participantes não é apenas um sinal de colaboração, mas também um indicador do nível de escrutínio e da potencial controvérsia de um PR, fatores que estão ligados a uma maior probabilidade de rejeição.

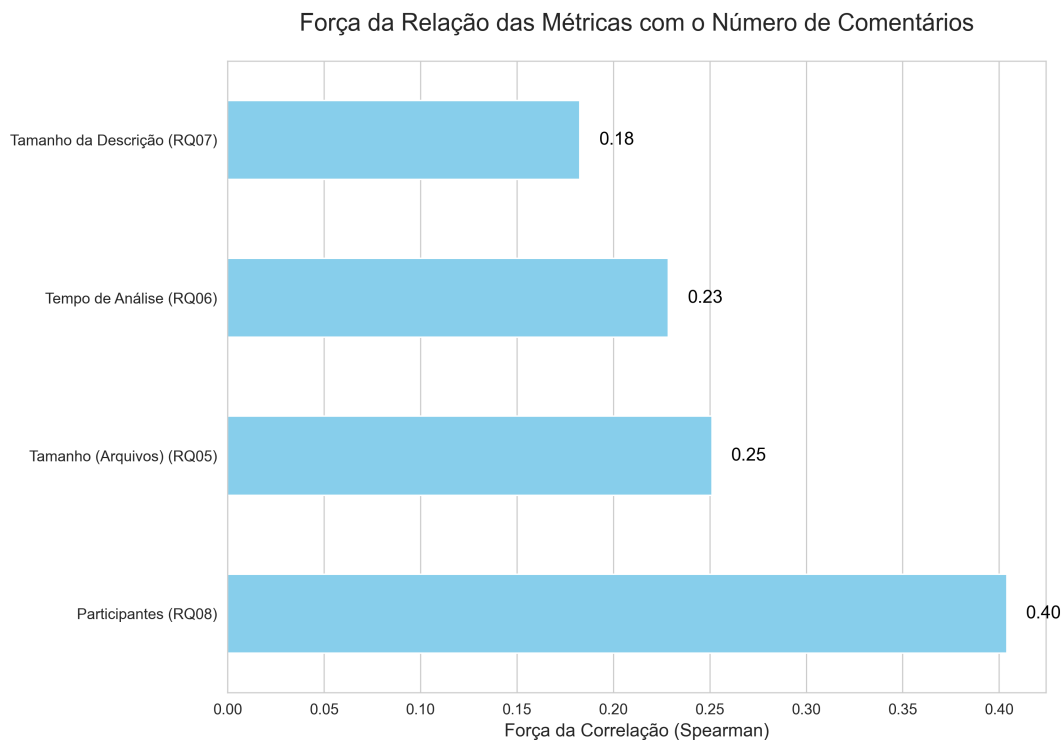


Figura 6: Força da correlação de Spearman entre as métricas e o número de comentários.

## 4 Conclusão

Este estudo se propôs a caracterizar a atividade de *code review* no GitHub, identificando variáveis que influenciam o desfecho dos Pull Requests. A análise empírica de um vasto conjunto de dados demonstrou que o sucesso de uma contribuição de código não depende apenas de sua qualidade técnica, mas está intrinsecamente ligado à dinâmica do processo de revisão e comunicação.

As conclusões podem ser sumarizadas em duas frentes principais:

1. **Fatores Preditivos para Aceitação de um PR:** A análise revelou que o perfil de um PR com alta probabilidade de ser aceite (MERGED) é claro e consistente. Tais contribuições são caracterizadas por serem de **menor escopo** (RQ01), possuírem **descrições detalhadas** (RQ03), serem resolvidas em um **tempo de análise curto** (RQ02) e gerarem um **volume baixo de interações**, indicando clareza e baixo atrito (RQ04).
2. **Fatores que Impulsionam o Esforço de Revisão:** O esforço de revisão, medido pelo volume

de comentários, é mais influenciado pela **complexidade social** da mudança. O **número de participantes** envolvidos na discussão (RQ08) é o preditor mais forte para um alto volume de comentários. A complexidade técnica, representada pelo **tamanho do PR** (RQ05) e pelo **tempo que ele permanece em aberto** (RQ06), também são fatores moderados que contribuem para um maior esforço de revisão.

O resultado central deste estudo reforça que a eficiência no desenvolvimento de software colaborativo depende de uma comunicação clara e da gestão proativa da complexidade. Para **desenvolvedores**, a implicação prática é a importância de submeter contribuições pequenas, focadas e muito bem documentadas para maximizar as chances de sucesso. Para os **mantenedores de projetos**, os dados ressaltam a necessidade de monitorar PRs que permanecem abertos por muito tempo ou que acumulam um volume excessivo de discussões, pois representam um alto risco de se tornarem esforço perdido.

Em suma, a qualidade de um Pull Request reside tanto no código quanto na clareza e agilidade do processo de sua revisão.