

Caracterizando a atividade de code review no GitHub

Lucas Cabral Soares
Lucas Hemétrio Teixeira
Maria Eduarda Amaral Muniz

1. Introdução

A prática de code review desempenha um papel crucial na garantia da qualidade do código e na redução de defeitos nos projetos de software, especialmente em ambientes open source como os hospedados no GitHub. Este laboratório visa investigar as variáveis que influenciam o feedback final e o número de revisões de Pull Requests (PRs) em repositórios populares do GitHub.

1.1. Hipóteses

Antes da análise dos dados, formulamos algumas hipóteses sobre os sistemas populares:

- **RQ01: Qual a relação entre o tamanho dos PRs e o feedback final das revisões?**
PRs com um número maior de arquivos ou com mais linhas modificadas tendem a receber um feedback final mais crítico ou com maior número de rejeições, devido à complexidade das mudanças. Acredita-se que PRs menores e mais focados possuam maior chance de aprovação rápida.
- **RQ02: Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?**
PRs que passam por um tempo de análise mais longo tendem a receber um feedback final mais detalhado e positivo, uma vez que as revisões mais cuidadosas e prolongadas resultam em menos defeitos e mais precisão nas alterações.
- **RQ03: Qual a relação entre a descrição dos PRs e o feedback final das revisões?**
PRs com descrições mais completas e detalhadas provavelmente recebem um feedback final mais favorável, pois fornecem um contexto melhor para os revisores e facilitam a compreensão das alterações propostas.
- **RQ04: Qual a relação entre as interações nos PRs e o feedback final das revisões?**

PRs que geram mais interações, como discussões e comentários entre os colaboradores, têm maior chance de obter um feedback final positivo, pois indicam que o código foi mais analisado e ajustado conforme as sugestões dos revisores.

- **RQ05: Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?**

PRs maiores em termos de arquivos modificados ou linhas alteradas exigem um maior número de revisões, já que as mudanças são mais complexas e precisam de mais checagens para garantir sua qualidade.

- **RQ06: Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?**

PRs que passam por uma análise mais prolongada tendem a passar por mais revisões, já que revisores podem fazer ajustes adicionais ao longo do tempo, à medida que identificam diferentes aspectos da modificação.

- **RQ07: Qual a relação entre a descrição dos PRs e o número de revisões realizadas?**

PRs com descrições mais completas tendem a gerar um maior número de revisões, pois ajudam os revisores a compreender melhor as intenções do autor, resultando em mais feedbacks e ajustes.

- **RQ08: Qual a relação entre as interações nos PRs e o número de revisões realizadas?**

PRs que envolvem um maior número de interações entre os colaboradores (comentários e discussões) tendem a resultar em um maior número de revisões, pois as discussões promovem ajustes constantes e mais detalhados no código proposto.

2. Metodologia

Para responder às questões de pesquisa, utilizamos uma abordagem baseada na coleta, processamento e análise de dados de repositórios populares do GitHub. A seleção dos repositórios foi feita considerando os 200 projetos com maior número de estrelas e que tenham no mínimo 100 PRs válidos, garantindo que nossa análise se concentrasse nos sistemas mais reconhecidos e amplamente utilizados na plataforma.

A metodologia adotada pode ser dividida em duas etapas principais: coleta dos dados, processamento das métricas e análise dos resultados.

2.1 Coleta dos Dados do GitHub

Os dados foram extraídos utilizando a GitHub GraphQL API, o que permitiu coletar informações detalhadas sobre cada repositório - as informações e métricas serão detalhadas na seção seguinte. Para garantir uma amostra representativa, utilizamos uma consulta paginada para coletar até 200 repositórios de forma eficiente. Os dados foram armazenados em um arquivo CSV para facilitar o processamento posterior.

2.2 Métricas

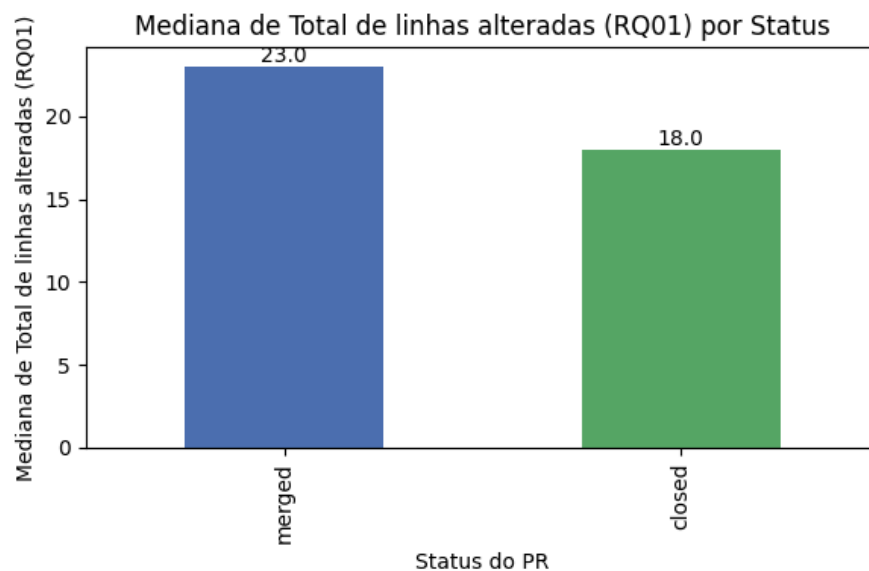
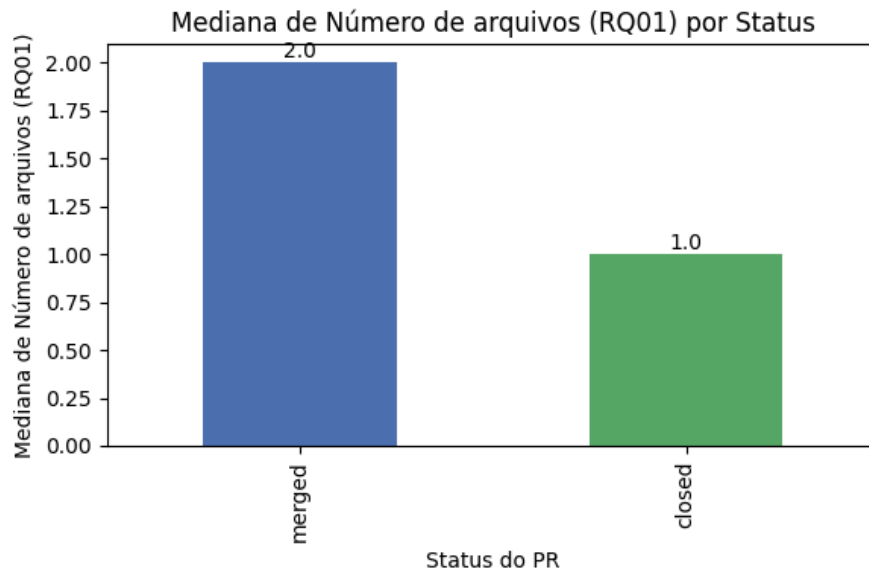
As métricas utilizadas para responder às questões de pesquisa foram as seguintes:

- **Tamanho:** total de linhas adicionadas e removidas
- **Tempo de Análise:** intervalo entre a criação do PR e a última atividade
- **Descrição:** número de caracteres do corpo de descrição do PR (na versão markdown)
- **Interações:** número de participantes e número de comentários
- **Status:** situação do PR, se foi concluído (merged) ou fechado (closed)
- **Reviews:** quantidade de reviews que o PR passou

3. Resultados e Discussão

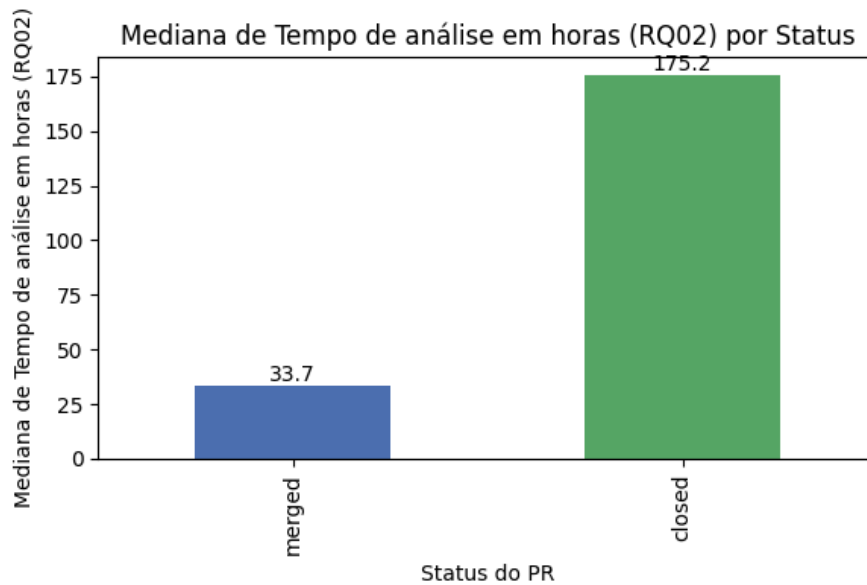
Todos os gráficos foram gerados com a biblioteca matplotlib e foram usados os valores medianos para a geração dos gráficos. Para os gráficos de correlação foi optado por usar Spearman pelo dataset tender a ter distribuições fortemente assimétricas, que tem um cálculo mais robusto para valores extremos.

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



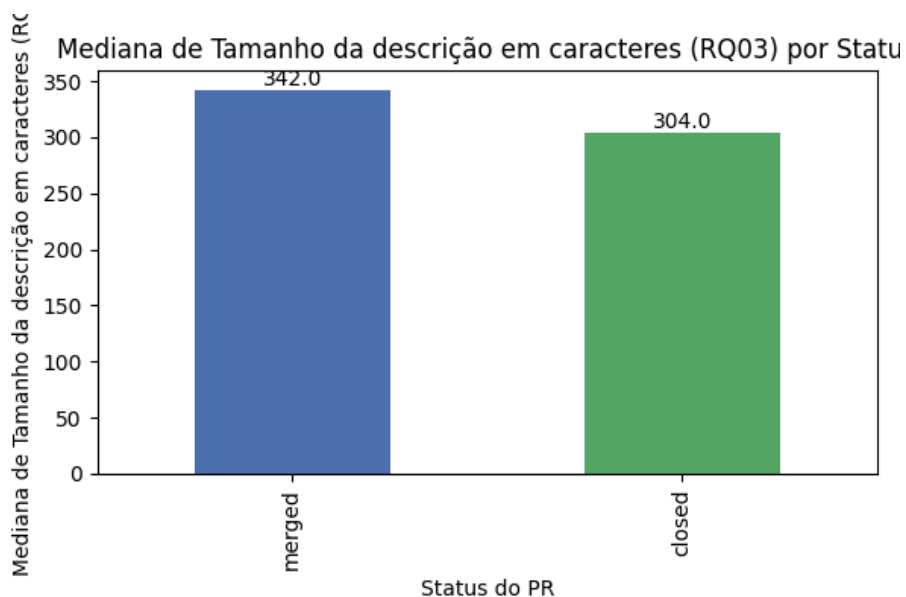
No gráfico de RQ1 observa-se que a mediana de arquivos modificados em PRs aprovados (2) supera a dos PRs fechados sem merge (1), assim como o total de linhas alteradas (23 versus 18). Esses resultados indicam que, ao contrário da hipótese de que mudanças mais extensas dificultariam a aprovação, PRs com maior escopo foram, de fato, integrados com mais frequência. Portanto, nossa suposição de que PRs menores teriam maior probabilidade de aprovação rápida não se confirma, levando à rejeição da hipótese.

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



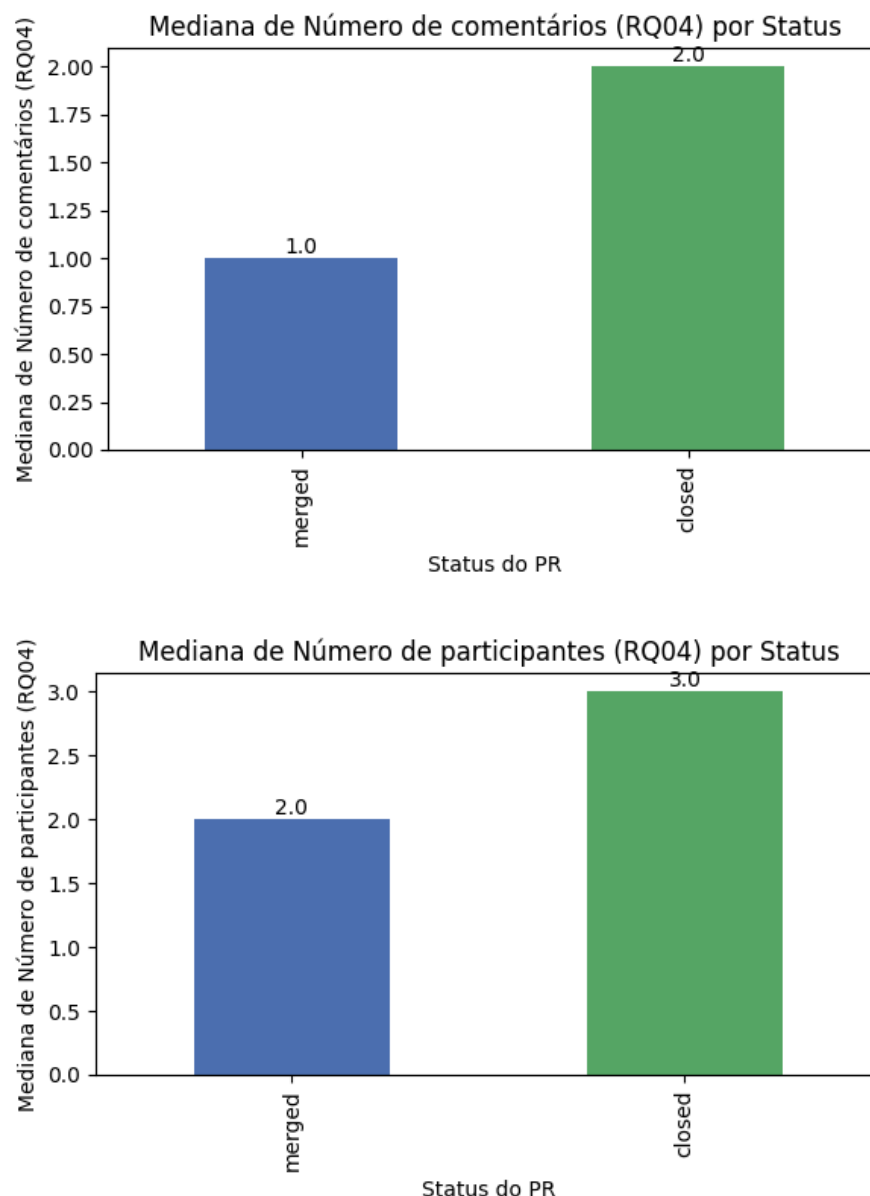
No gráfico de RQ2 verifica-se que a mediana de tempo de análise para PRs aprovados (33,7 h) é muito inferior à dos PRs fechados sem merge (175,2 h), indicando que quanto mais longa é a janela de revisão, maior a probabilidade de o PR acabar sendo rejeitado ou encerrado sem merge. Esse resultado contradiz a hipótese de que análises mais demoradas resultariam em feedback mais positivo, pois aqui são justamente os PRs de revisão rápida que chegam ao merge. Logo, a hipótese de que maior tempo de análise favorece aprovações é rejeitada.

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



No gráfico de RQ3 observa-se que a mediana do tamanho da descrição em caracteres para PRs aprovados (342) é superior àquela dos PRs fechados sem merge (304), indicando que descrições mais longas estão associadas a feedbacks finais mais favoráveis; portanto, nossa hipótese de que PRs com descrições mais completas e detalhadas recebem aprovação com mais frequência fica validada.

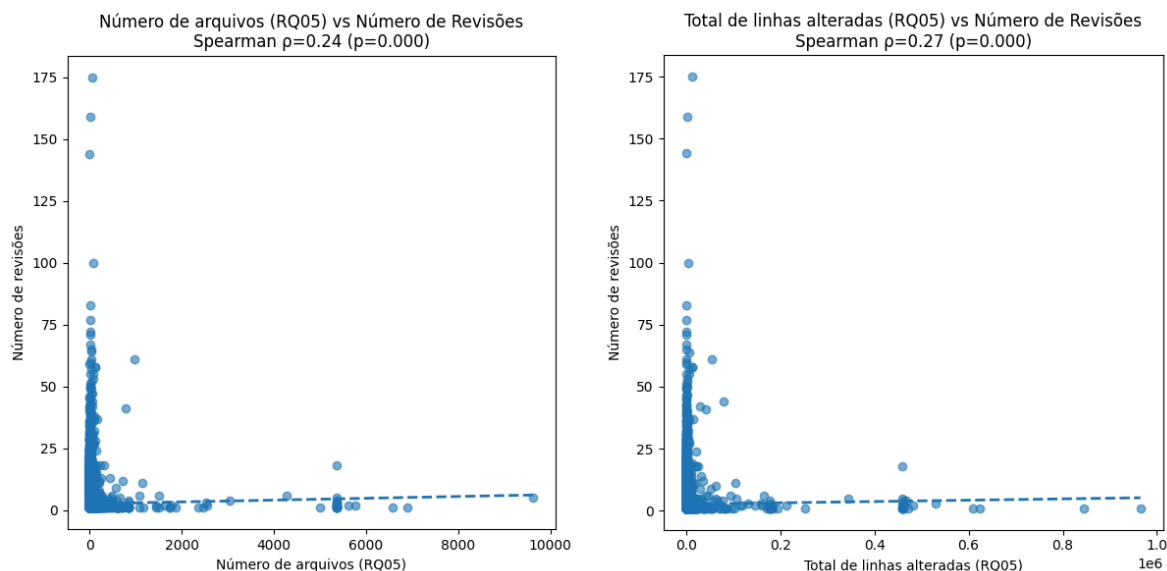
3.4 RQ 04. Qual a relação entre as interações nos PRs e o feedback final das revisões?



No gráfico de RQ04 observa-se que os PRs fechados sem merge apresentam mediana maior tanto de participantes (3 versus 2) quanto de comentários (2 versus 1) em comparação aos PRs aprovados, sugerindo que as revisões que geram mais interações não resultam em feedback final mais favorável, mas sim em

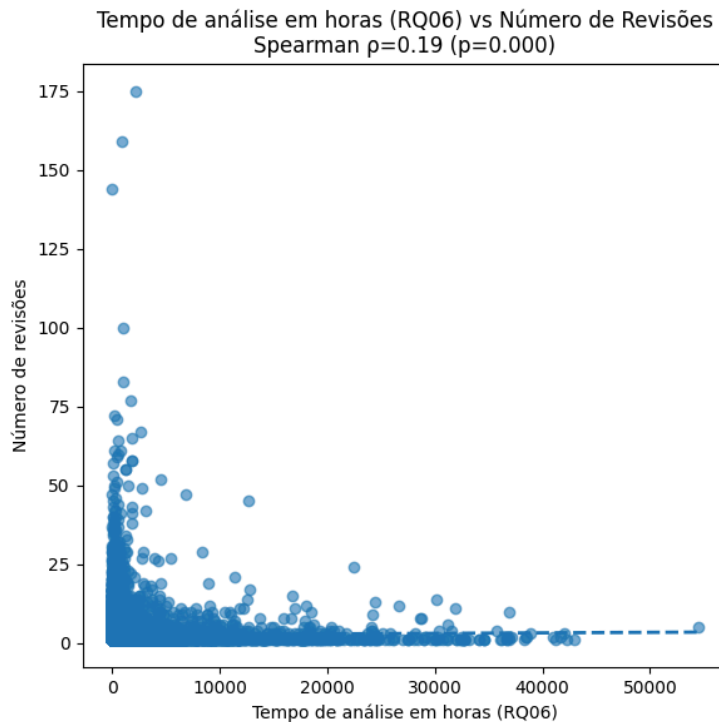
encerramento sem merge. Dessa forma, a hipótese de que PRs com mais discussões e comentários tenderiam a ser aprovados é rejeitada.

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



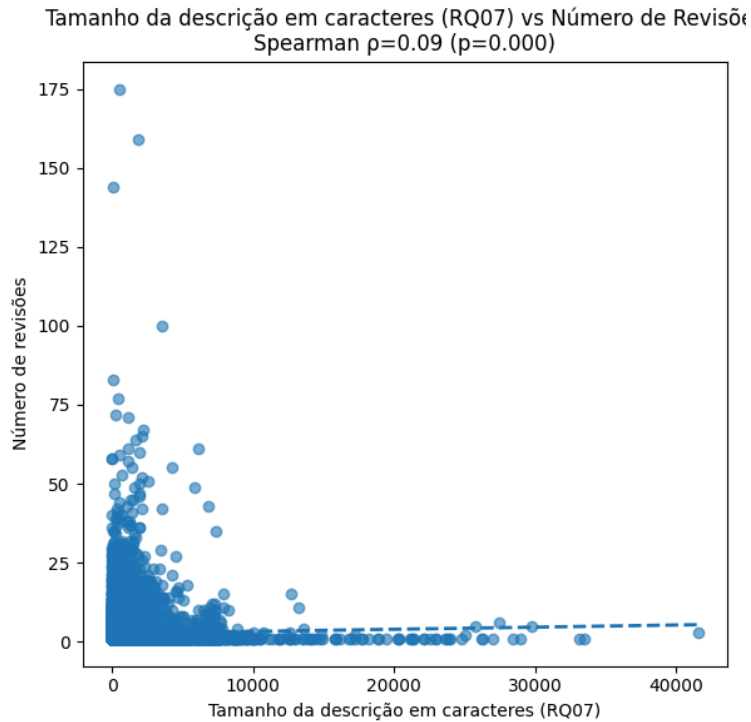
No gráfico de RQ05 observa-se uma correlação positiva, embora fraca, entre o tamanho dos PRs (tanto em total de linhas alteradas quanto em número de arquivos) e o número de revisões, com Spearman $\rho = 0,27$ ($p < 0,001$) para linhas e $\rho = 0,24$ ($p < 0,001$) para arquivos, e as linhas de tendência destacam esse leve aumento no número de revisões à medida que crescem as mudanças. Esses resultados indicam que PRs maiores tendem, de fato, a demandar mais checagens, mas o efeito é modesto, sugerindo que outros fatores também influenciam fortemente a quantidade de revisões. Assim, nossa hipótese é parcialmente validada.

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



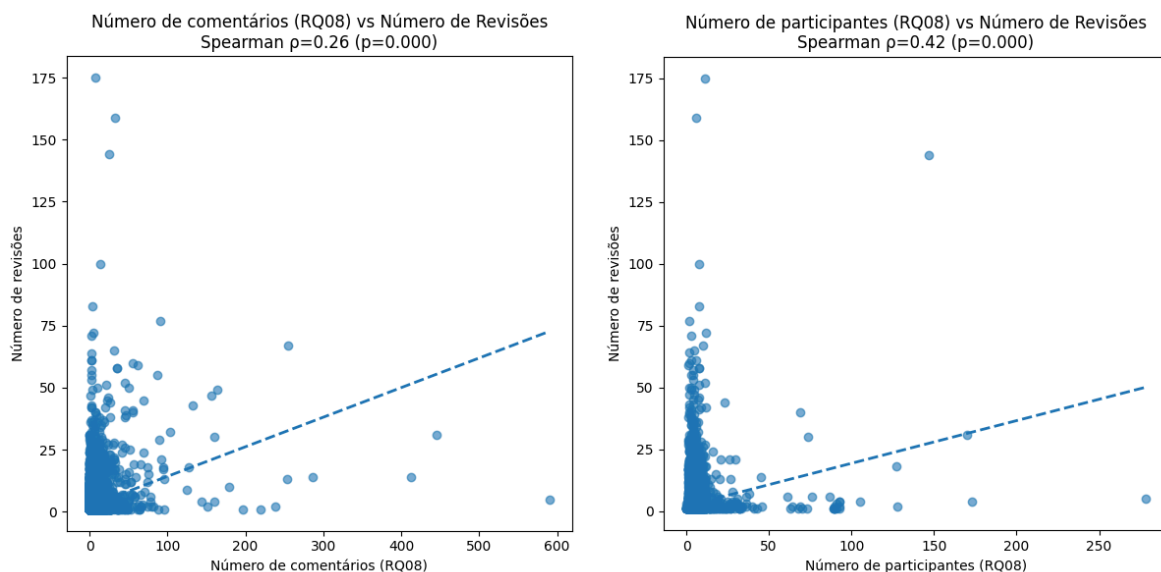
No gráfico de RQ06 observa-se uma correlação positiva, embora fraca, entre o tempo de análise e o número de revisões (Spearman $\rho = 0,19$; $p < 0,001$): a linha de tendência mostra um leve aumento nas revisões conforme o tempo de análise cresce, mas a dispersão dos pontos indica que esse efeito é modesto. Dessa forma, confirma-se que PRs com janelas de revisão mais longas tendem a passar por mais ciclos de revisão, mas o impacto não é forte, de modo que nossa hipótese fica apenas parcialmente validada.

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



No gráfico de RQ07 observa-se que o tamanho da descrição em caracteres apresenta uma correlação positiva, porém muito fraca, com o número de revisões (Spearman $\rho = 0,09$; $p < 0,001$), e a linha de tendência aponta apenas um ligeiro aumento nas revisões à medida que crescem as descrições. A ampla dispersão dos pontos revela que, na prática, descrições mais longas não implicam em acréscimos significativos de ciclos de revisão. Dessa forma, nossa hipótese de que descrições mais completas gerariam um número maior de revisões é apenas parcialmente validada.

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



No gráfico de RQ08 observa-se uma correlação moderada e positiva entre número de participantes e número de revisões (Spearman $\rho = 0,42$; $p < 0,001$) e uma correlação mais fraca, mas ainda significativa, entre número de comentários e revisões ($\rho = 0,26$; $p < 0,001$). Esses resultados mostram que PRs que envolvem mais interações — tanto em termos de colaboradores distintos quanto de comentários — tendem realmente a passar por mais ciclos de revisão. Logo, nossa hipótese de que maior engajamento gera maior número de revisões é validada.

4. Conclusão

Em síntese, nossas análises revelam que nem sempre os padrões esperados se confirmam: PRs maiores — com mais arquivos e linhas alteradas — tendem a ser integrados com mais frequência e de forma mais rápida do que PRs menores, contrariando a ideia de que mudanças extensas sofram mais rejeições (RQ01 e RQ02). Por outro lado, descrições mais completas mostram efeito positivo sobre a aprovação (RQ03) e, embora PRs com mais interações (comentários e participantes) apresentem mais ciclos de revisão, essa mesma intensidade de feedback acaba associada a um maior índice de fechamento sem merge (RQ04).

Quando investigamos o número de revisões, verificamos correlações positivas porém moderadas: mudanças de maior escopo, janelas de análise mais longas e descrições mais detalhadas geram um aumento discreto no volume de revisões (RQ05–RQ07), enquanto a quantidade de colaboradores envolvidos e de comentários exerce influência mais forte sobre o número de revisões realizadas (RQ08). Esses resultados sugerem que, para além do tamanho e da complexidade técnica, o engajamento social no code review é o principal motor de múltiplos ciclos de revisão, destacando a importância de processos colaborativos bem estruturados.