



PUC Minas

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Instituto de Ciências Exatas e Informática — Unidade Praça da Liberdade

Curso: Engenharia de Software.

Prof: José Laerte Pires Xavier Junior

Disciplina: Laboratório de Experimentação de Software

Alunos: Bruno Henrique Armanelli e Lucas Alves Gusmão

Características de Qualidade de Sistemas Java

1. Introdução

As métricas CK são um conjunto de métricas utilizadas para medir características de projetos escritos em linguagens orientadas a objetos.

O GitHub é a maior coleção de projetos *open-source* na internet. A plataforma é utilizada tanto por desenvolvedores que desejam publicar seus projetos pessoais de pequeno porte, quanto por empresas e organizações que disponibilizam projetos e ferramentas para a comunidade. Por conta disso, ele torna-se uma fonte interessante para análises e pesquisas relacionadas à área da engenharia de software, sendo possível tirar conclusões com base nos dados armazenados na plataforma.

Com base em análises de dados por meio da API pública do GitHub, foi feito um trabalho na disciplina de Laboratório de Medição e Experimentação de Software com o objetivo de definir e verificar hipóteses sobre repositórios escritos na linguagem Java que são populares na plataforma, com base nas métricas CK. A popularidade será definida com base no total de "estrelas" que cada repositório recebeu de usuários na plataforma.

2. Metodologia

Este trabalho é uma pesquisa descritiva com base em uma abordagem quantitativa sob a base de dados pública do GitHub. Os dados foram coletados através de consultas na API em GraphQL disponibilizada pela plataforma. Na consulta foram limitados os primeiros 1000 últimos repositórios que atendem à regra de ordenação por estrelas e utilizam a linguagem. Um pequeno programa foi escrito na linguagem JavaScript, utilizando a biblioteca Apollo como cliente GraphQL, para executar a consulta e percorrer todas as páginas na API. Em seguida, o programa realizou clone de cada repositório e executou a ferramenta CK para a coleta e armazenamento de resultado das métricas produzidas por ela. Finalmente, os dados foram extraídos para um arquivo no formato CSV e importados para a ferramenta DataStudio do Google, para que pudessem ser analisados.

As RQs, assim como suas métricas e hipóteses são as seguintes:

- **RQ 01. Qual a relação entre a popularidade dos repositórios e as suas características de qualidade?**

Métricas: número de estrelas, CBO, DIT, LCOM

Hipótese: a hipótese levantada é de que quanto maior a popularidade (número de estrelas), menores serão os valores dos resultados obtidos na análise de CBO, DIT e LCOM, por se tratarem de projetos grandes e com muitos contribuidores.

- **RQ 02. Qual a relação entre a *maturidade* dos repositórios e as suas características de qualidade?**

Métrica: idade (em anos), CBO, DIT, LCOM

Hipótese: a hipótese levantada é de que a idade não influencia nos valores obtidos nas análises de CBO, DIT e LCOM, porque a idade não tem relação com a qualidade do repositório.

- **RQ 03. Qual a relação entre a atividade dos repositórios e as suas características de qualidade?**

Métrica: número de releases, CBO, DIT, LCOM

Hipóteses: é esperado na hipótese de relação com número de releases de que os repositórios com a maior quantidade de releases tendem a ter um maior número de LCOM, e menores números de CBO e LCOM, por razão de repositórios com maiores números de releases passarem por mais revisões de código.

- **RQ 04. Qual a relação entre o tamanho dos repositórios e as suas características de qualidade?**

Métrica: LOC, CBO, DIT, LCOM

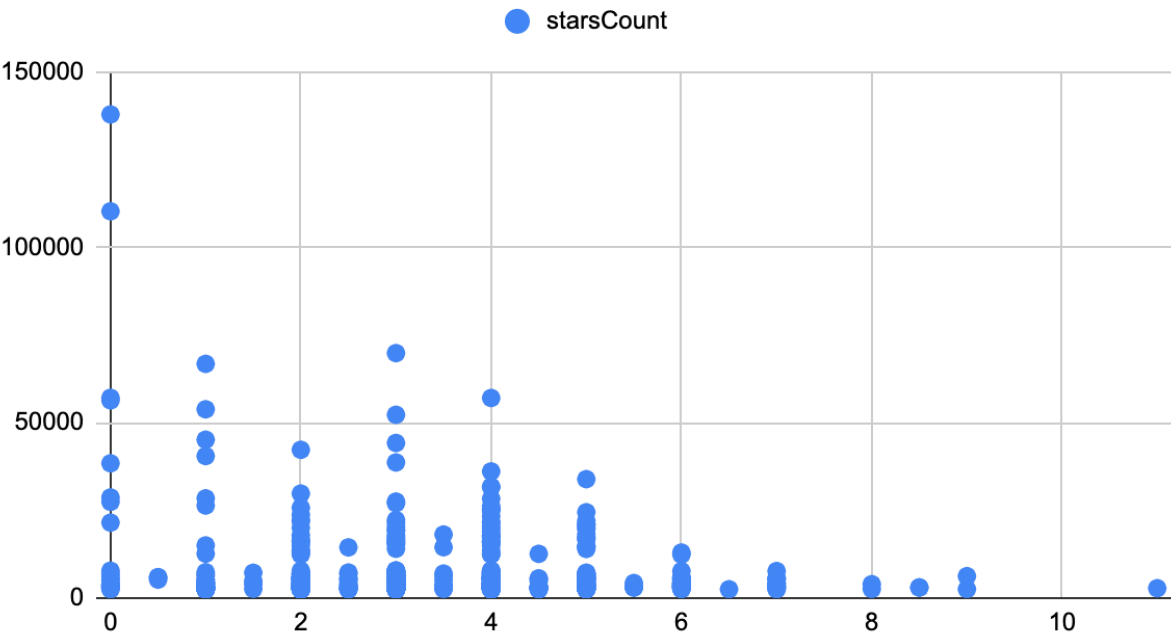
Hipótese: a hipótese esperada é de que os repositórios de maior tamanho em linhas de código tenham menores valores de LCOM e ao mesmo tempo maior número de CBO, com o DIT na maior parte maior que zero.

3. Resultados Obtidos

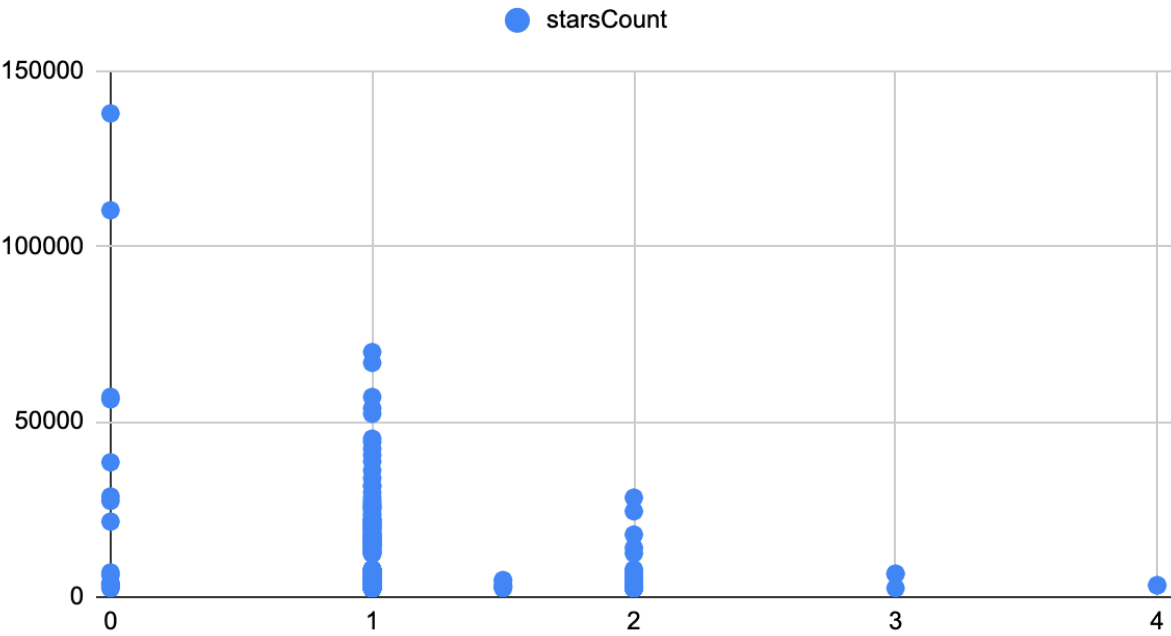
Com os dados obtidos na execução do script, foram elaboradas respostas com os resultados para cada *Research Question*. A execução e coleta dos dados ocorreu em 12 de agosto de 2021.

- **RQ 01:** foram elaborados os seguintes gráficos para demonstrar a correlação entre popularidade e qualidade de um repositório:

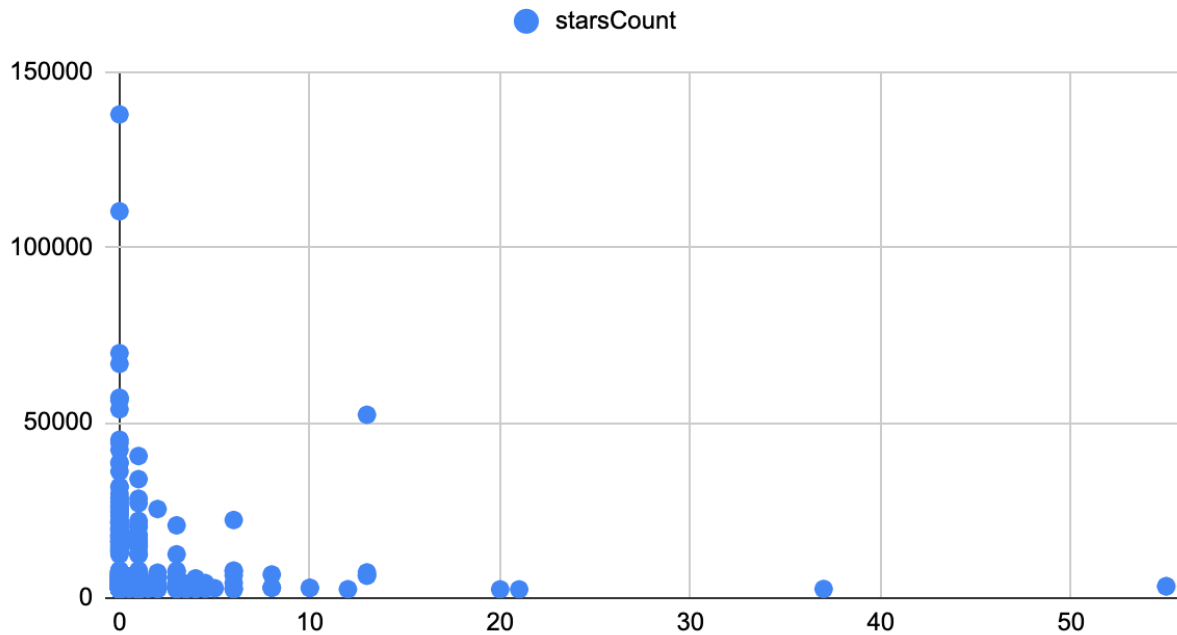
CBO / Quant. de Estrelas



DIT / Quant. de Estrelas



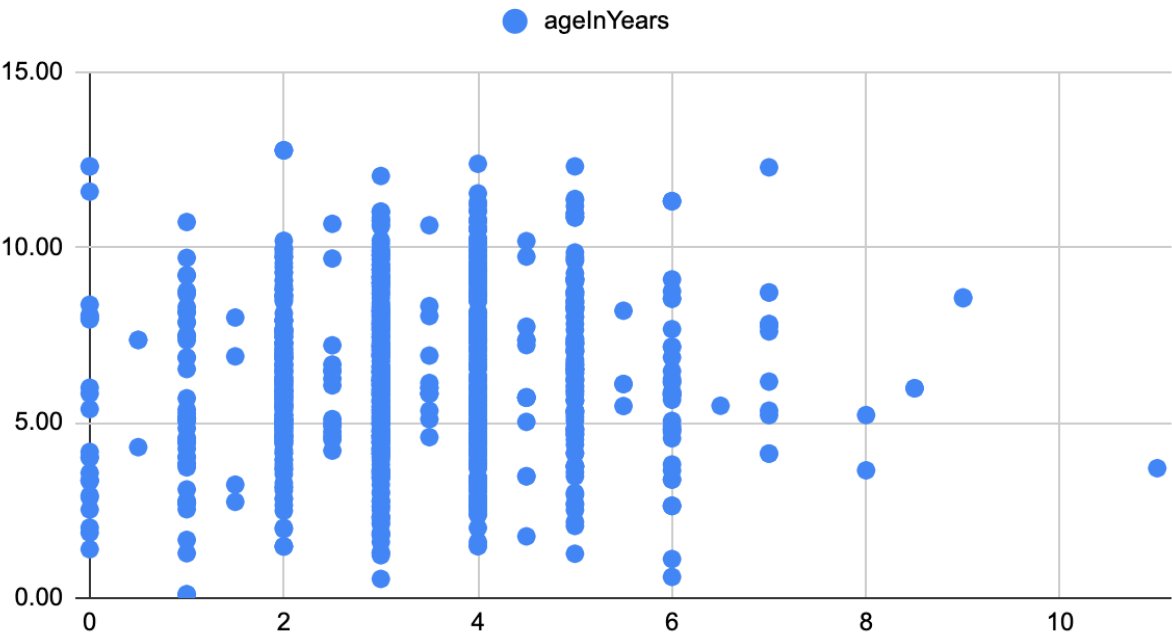
LCOM / Quant. de Estrelas



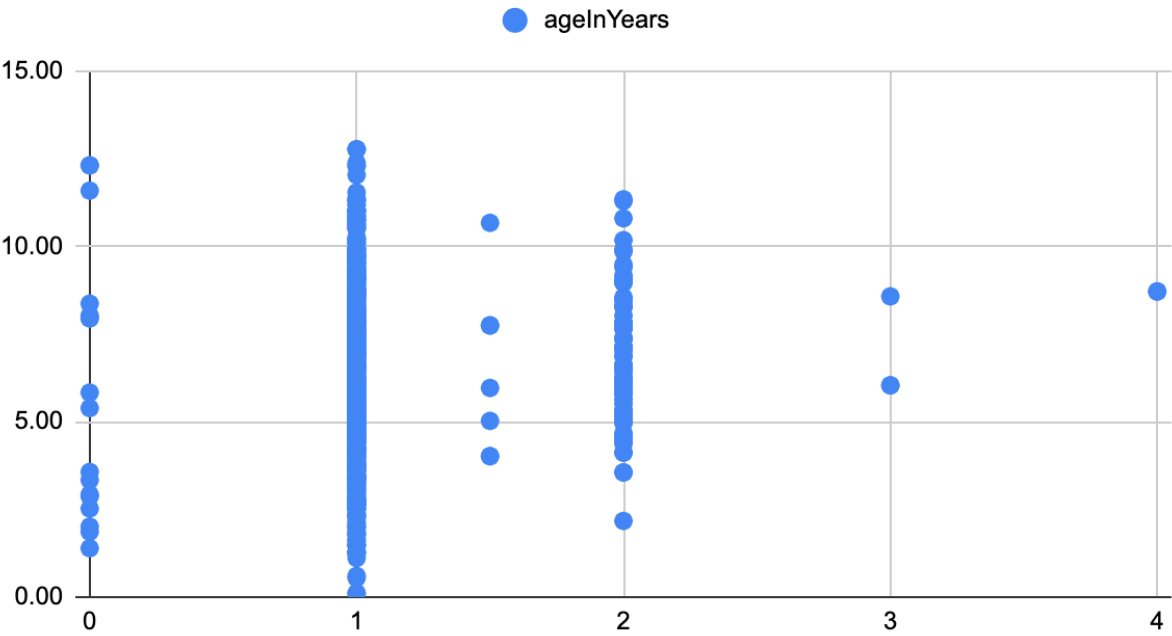
Foi possível perceber uma correlação entre a popularidade e as métricas de qualidade dos repositórios. A maioria dos repositórios que possuem números mais elevados de estrelas possui números menores de CBO, DIT e LCOM.

- **RQ 02:** foram elaborados os seguintes gráficos para demonstrar a correlação entre maturidade e qualidade de um repositório:

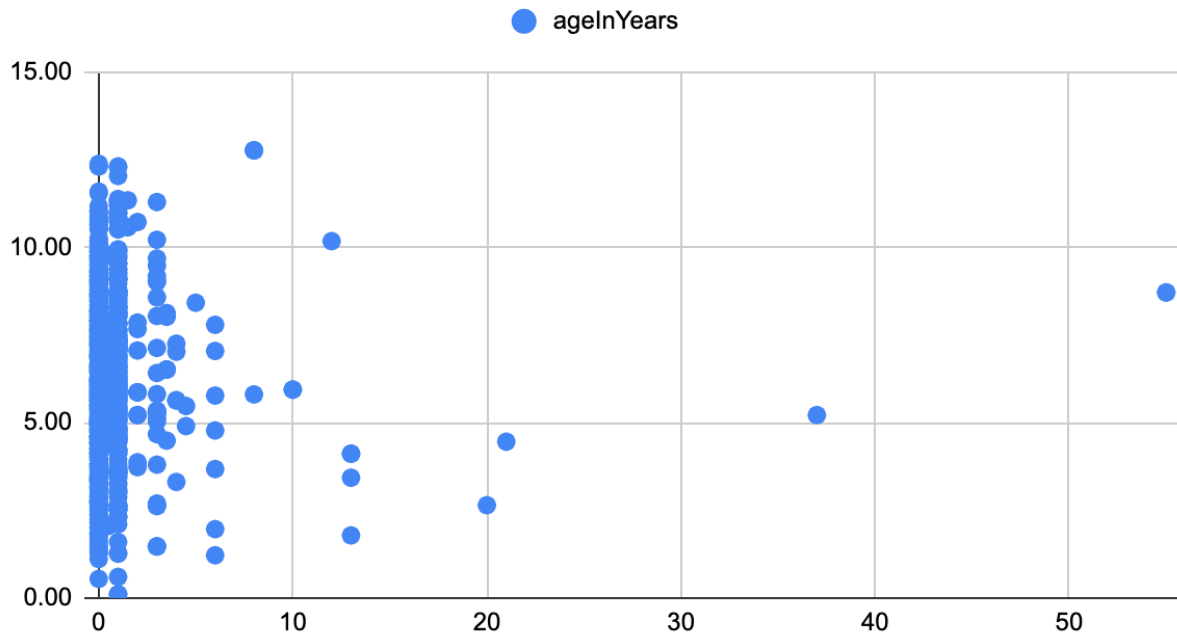
CBO / Idade do Repositório



DIT / Idade do Repositório



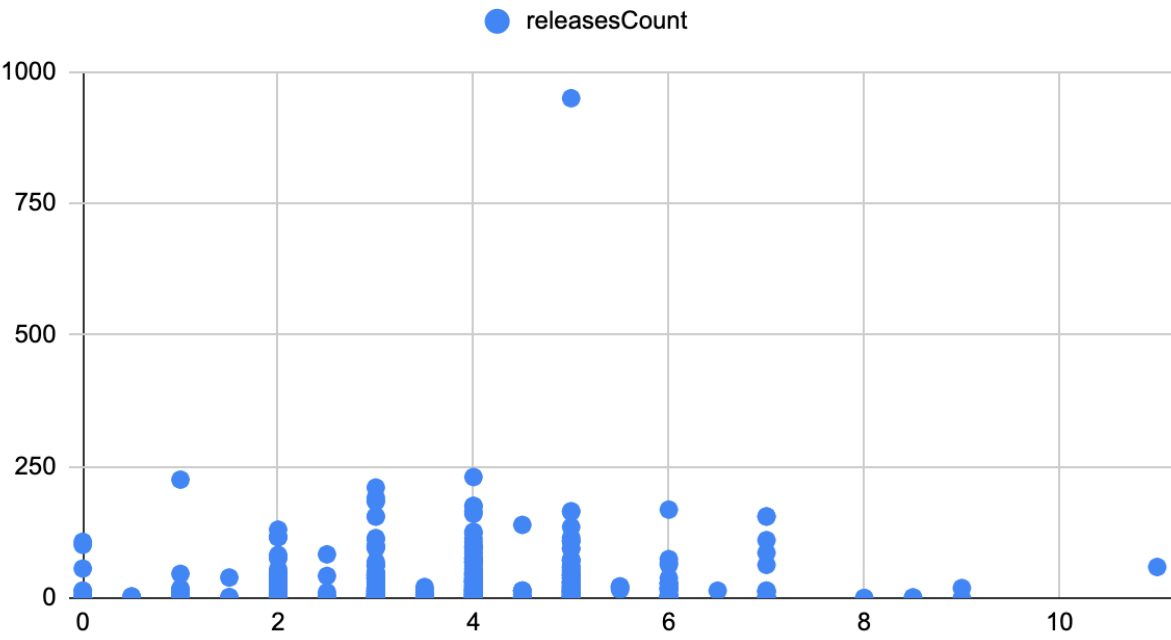
LCOM / Idade do Repositório



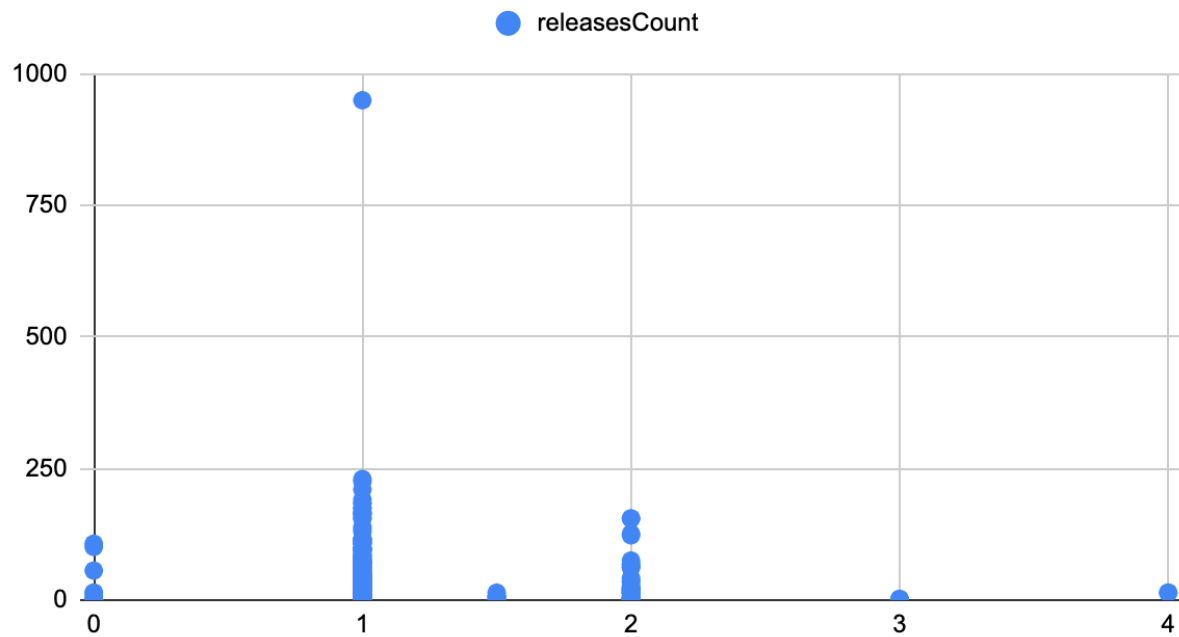
Não foi possível perceber uma correlação entre a maturidade e as métricas de qualidade dos repositórios. Existem repositórios de todas as idades nas mesmas faixas de resultados das métricas de qualidade, portanto, não foi possível identificar nenhuma tendência.

- **RQ 03:** foram elaborados os seguintes gráficos para demonstrar a correlação entre atividade e qualidade de um repositório:

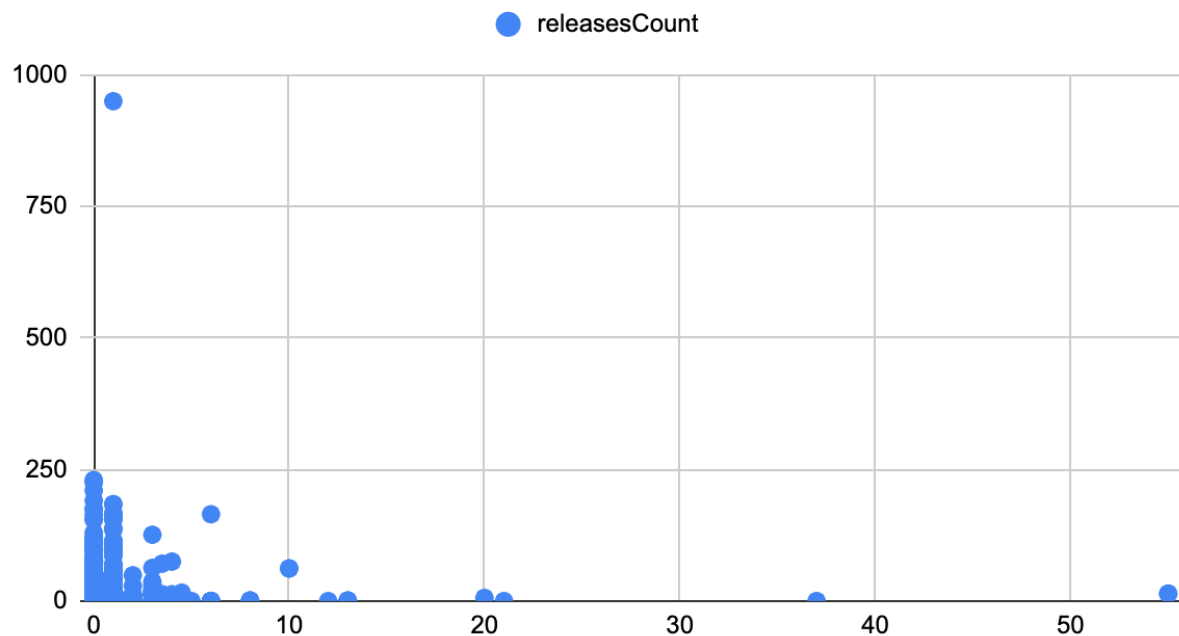
CBO / Quant. de Releases



DIT / Quant. de Releases



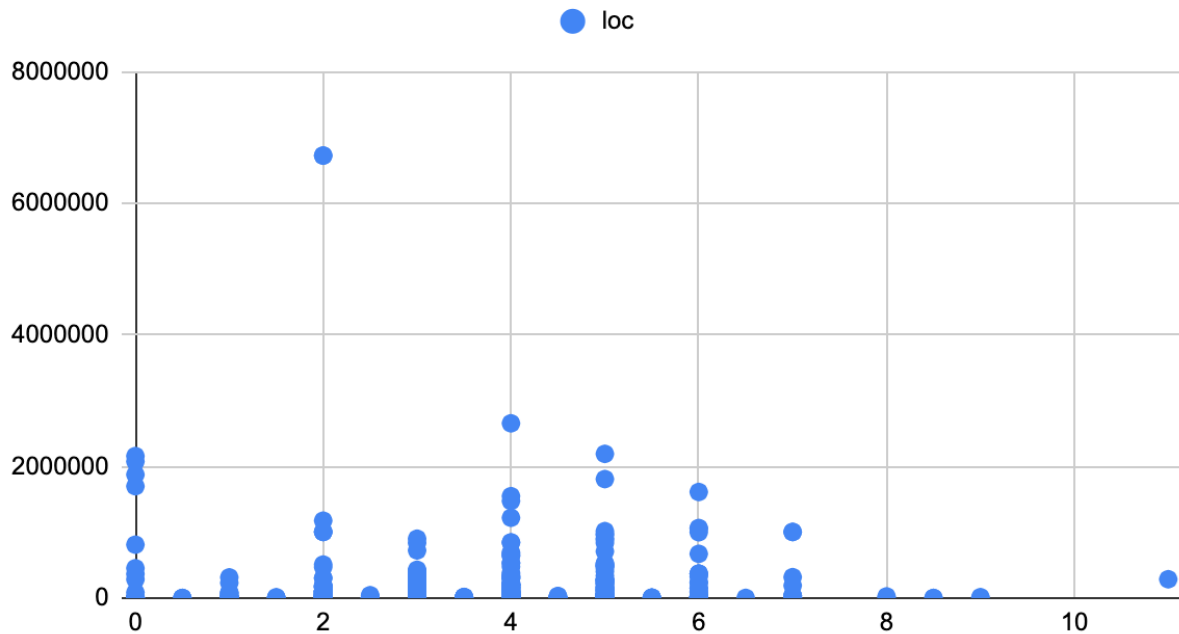
LCOM / Quant. de Releases



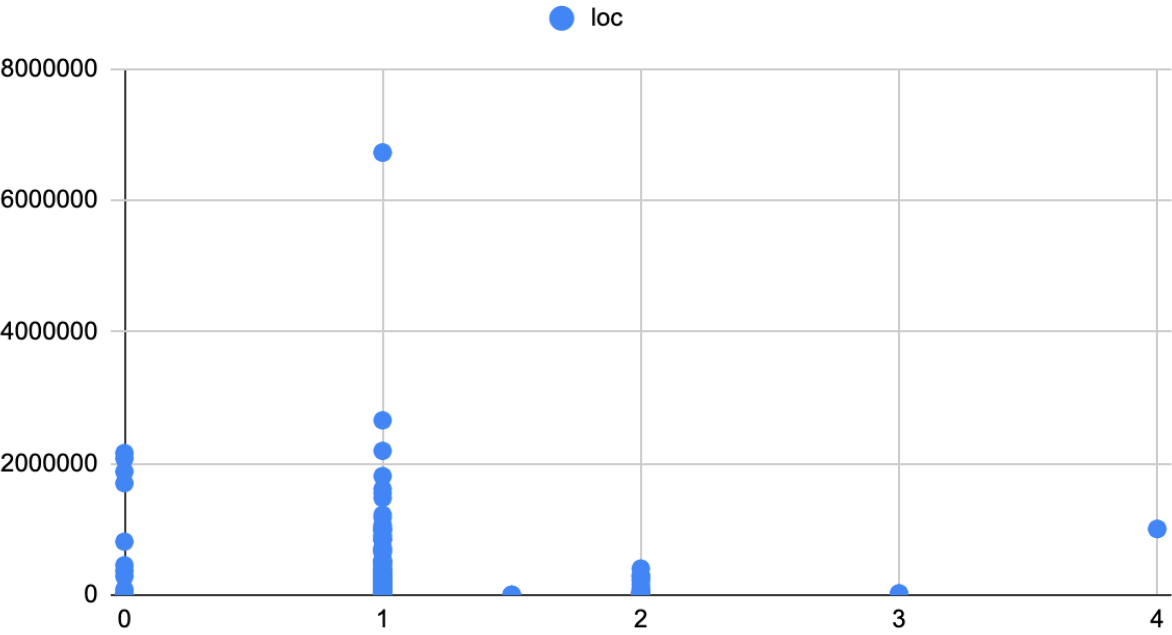
ao fato de que o recurso de releases não é utilizado por muitos projetos do GitHub.

- **RQ 04:** foram elaborados os seguintes gráficos para demonstrar a correlação entre maturidade e qualidade de um repositório:

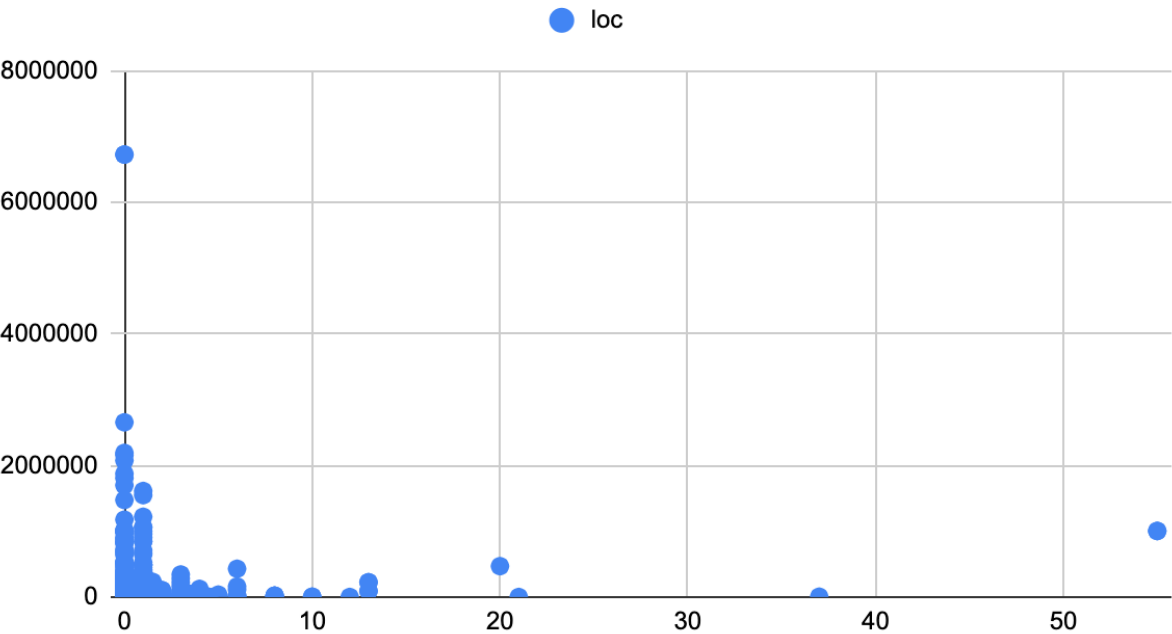
CBO / LOC



DIT / LOC



LCOM / LOC



Foi possível identificar uma correlação entre o número de linhas e as métricas de qualidade. A maioria dos repositórios possuem CBO inferior a 8 e DIT inferior a 2. Além disso, a maioria dos repositórios com quantidades mais elevadas de LOC possuíam LCOM abaixo de 10.

4. Conclusão

Considerando os resultados obtidos e as hipóteses levantadas, pode-se concluir, para cada pergunta, respectivamente:

- **RQ 01:** Foi definido na hipótese levantada anteriormente que haveria correlação positiva entre a popularidade e a qualidade de um repositório e essa correlação foi confirmada com os dados
- **RQ 02:** Foi levantada na hipótese anteriormente que não haveria correlação entre a maturidade e a qualidade de um repositório e essa falta de correlação foi confirmada com os dados
- **RQ 03:** Na hipótese foi especulado anteriormente que haveria uma correlação positiva nas métricas de CBO e DIT e negativa de LCOM entre a atividade e a qualidade de um repositório e essa correlação e essa correlação foi comprovada com os dados.
- **RQ 04:** A hipótese levantada em relação a esta RQ foi de que os repositórios com mais LOCs teriam menores valores de LCOM e maiores valores de CBO e DIT. Foi identificado com os dados coletados que existia uma correlação positiva entre as métricas de qualidade e o tamanho de um projeto.