

LABORATÓRIO DE EXPERIMENTAÇÃO DE SOFTWARE
UM ESTUDO DAS CARACTERÍSTICAS DE QUALIDADE DE SISTEMAS JAVA

Grupo: Helen Andrade, João Pedro, Kleyann Martins

Pontifícia Universidade Católica de Minas Gerais (PUC Minas)

Belo Horizonte - MG - Brasil

1. Introdução

O Objetivo do estudo realizado é analisar aspectos de qualidade nos repositórios mais populares na linguagem Java, hospedados no GitHub. Sendo utilizados como parâmetro métricas de produto e de qualidade calculadas através da ferramenta CK.

2. Hipóteses informais

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

Sistemas que possuem grande quantidade de contribuições possuem código limpo e bem documentado, incentivando a contribuição.

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

Sistemas mais maduros tiveram um maior tempo para evoluir e refinar os pontos necessários, e por isso possuem boas características de qualidade.

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

Sistemas com maior número de atividades (releases) apresentam uma maior qualidade devido às constantes melhorias e atualizações.

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

Repositórios com um tamanho moderado apresentam maior qualidade, já que repositórios muito grandes podem ser difíceis de manter e repositórios muito pequenos podem não possuir recursos suficientes para ter uma alta qualidade.

3. Metodologia

Este estudo selecionou os 1000 repositórios Java mais populares do GitHub baseando a popularidade pelo número de estrelas. Por meio destes repositórios foram obtidas as informações de data da criação, número de releases, nome do repositório e URL.

A coleta dos dados foi realizada mediante o uso da API GraphQL do GitHub e um script Python. Os dados foram salvos em um arquivo CSV para posterior análise.

Para analisar o código de cada um dos 1000 repositórios, foi realizada uma clonagem usando a URL coletada e a ferramenta CK foi executada para a obtenção das métricas de qualidade, sendo elas CBO, DIT e LCOM. Os dados obtidos pelo CK foram sumarizados para fins de análise.

Todos dados coletados foram submetidos a uma análise de correlação simples, de LOC com CBO, DIT e LCOM, estrelas com CBO, DIT e LCOM, idade com CBO, DIT e LCOM e releases com CBO, DIT e LCOM.

Para a análise dos resultados obtidos, os dados que continham valores vazios de LOC, CBO, DIT e LCOM foram removidos. Os resultados finais foram utilizados para gerar gráficos de dispersão criados por meio de um script Python que utilizou as bibliotecas Matplotlib e Pandas.

4. Resultados

Após realizar a análise dos dados coletados, é possível obter insights correlacionando características do processo de desenvolvimento com métricas de produto coletadas por meio da ferramenta CK.

4.1 Popularidade

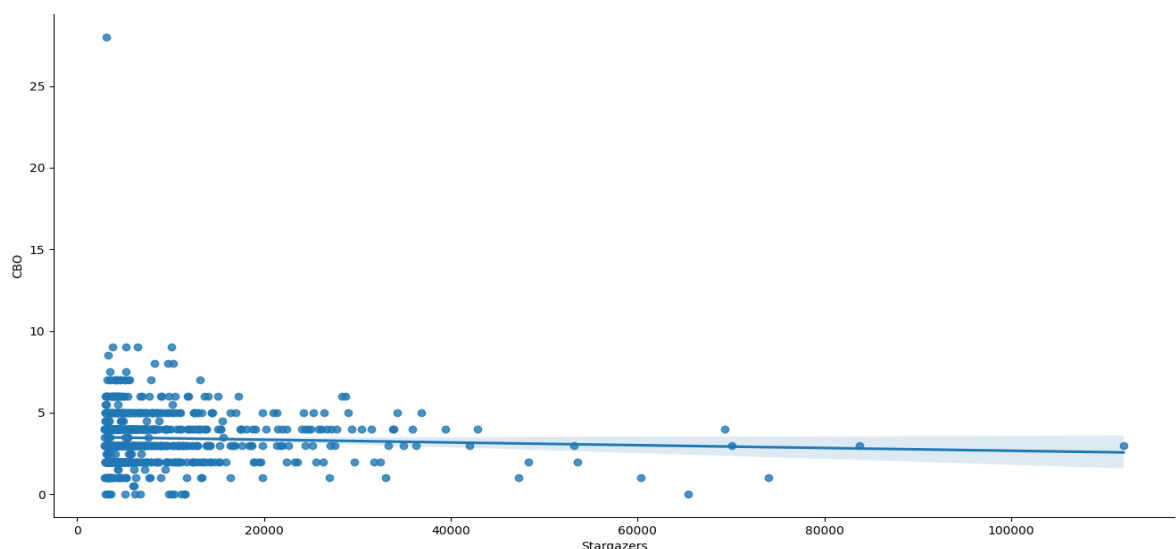


Figura 1. Popularidade dos repositórios correlacionados com a métrica de CBO.

Na Figura 1, pode-se observar uma fraca correlação entre a popularidade e o acoplamento entre objetos (CBO).

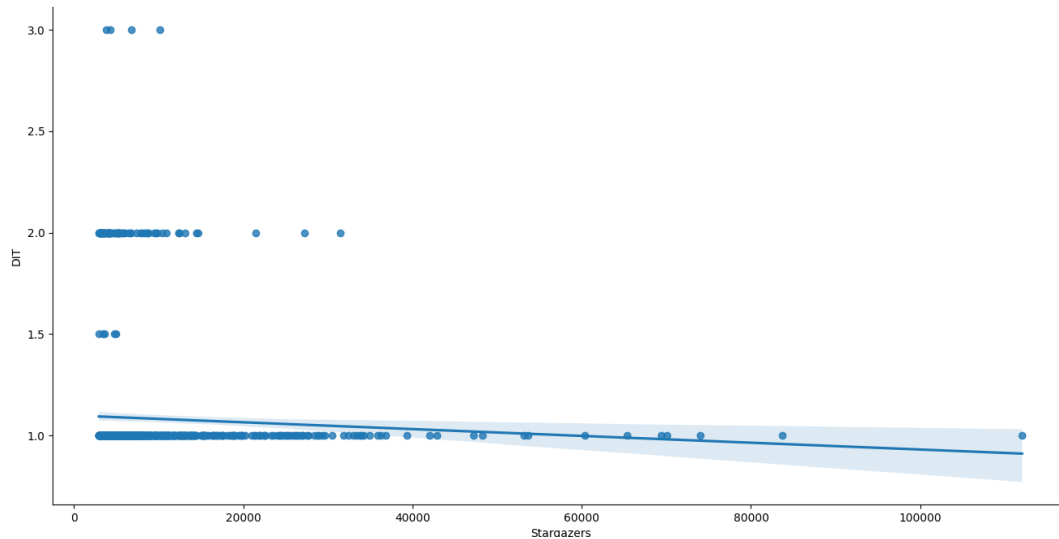


Figura 2. Popularidade dos repositórios correlacionados com a métrica de DIT.

Na Figura 2 é possível identificar a fraca correlação entre a popularidade dos repositórios e a métrica DIT.

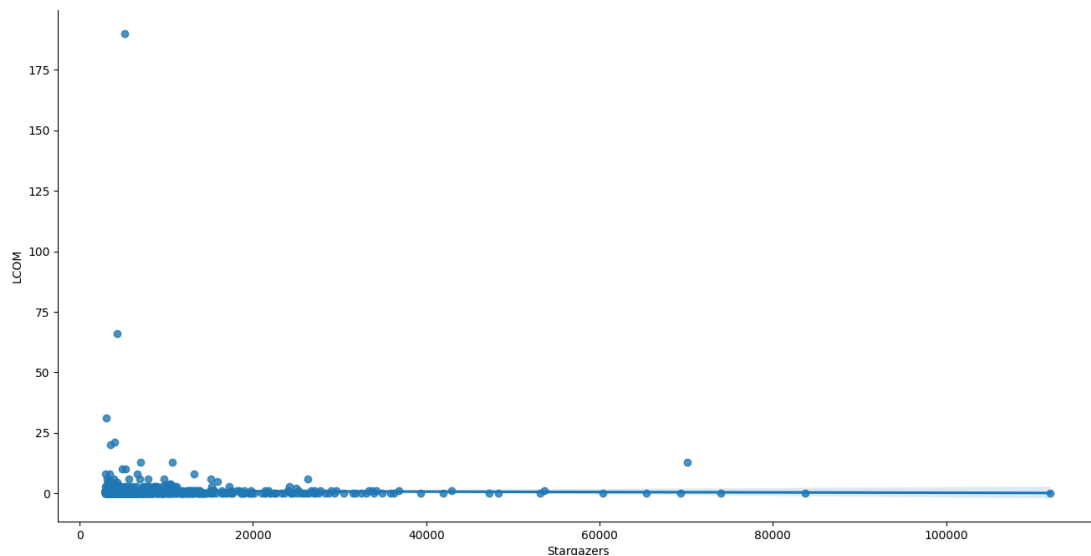


Figura 3. Popularidade dos repositórios correlacionados com a métrica de LCOM.

Na Figura 3, existe uma fraca correlação entre a popularidade dos repositórios e o LCOM. Tendo os dados bem próximos de 0, a partir dos dados analisados é possível concluir que a popularidade não exerce influência sobre a falta de coesão.

4.2 Maturidade

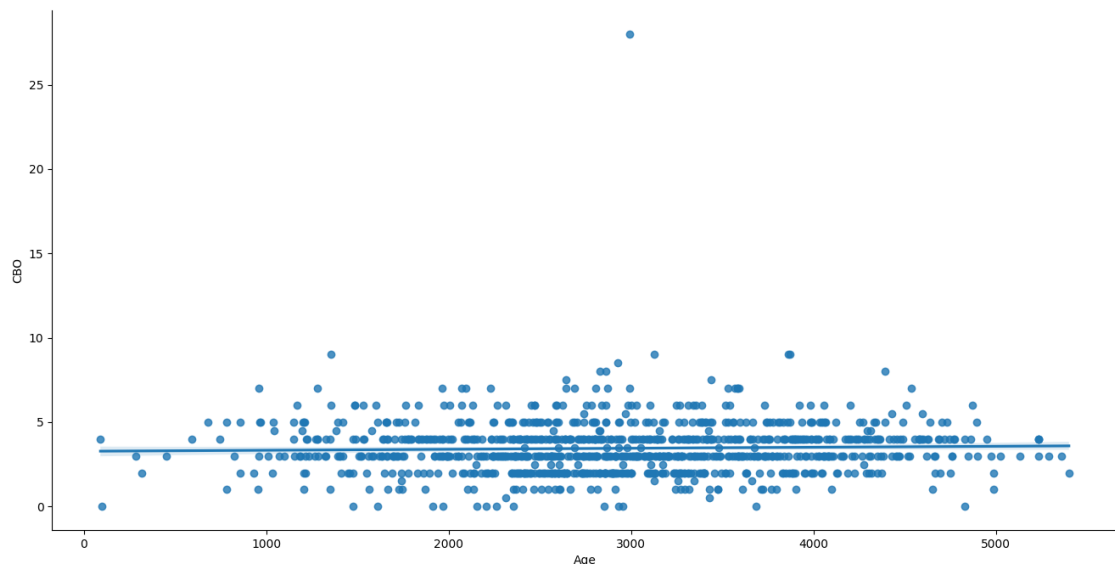


Figura 5. Idade dos repositórios correlacionada com as métricas do CBO.

Na Figura 5, pode-se identificar uma fraca correlação entre a maturidade dos repositórios e o acoplamento entre objetos (CBO).

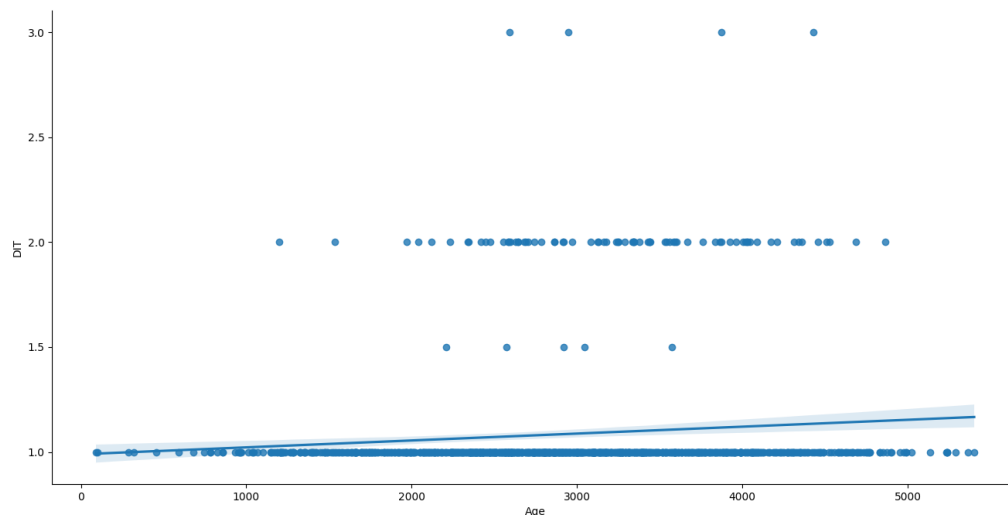


Figura 6. Idade dos repositórios correlacionada com a métrica DIT.

Na Figura 6 é possível observar uma fraca correlação entre a maturidade dos repositórios e o DIT. Tendo a maior parte dos valores muito próximos de 0, sendo possível concluir que a popularidade não influencia no número de níveis que uma classe herda.

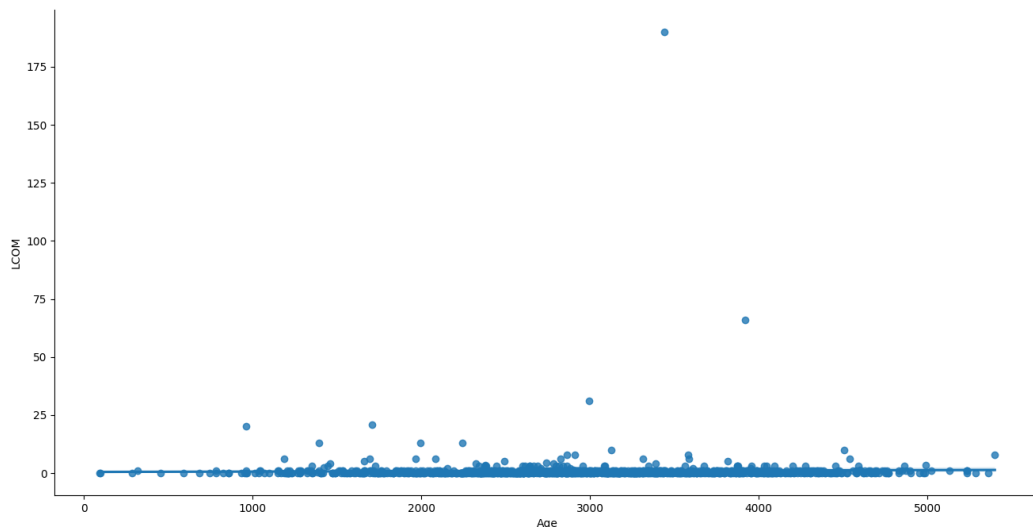


Figura 7. Idade dos repositórios correlacionada com a métrica LCOM.

Na Figura 7 existe uma fraca relação entre a idade dos repositórios e a métrica de LCOM. A partir dos dados tendo a maior parte dos valores próximos de 0, é possível concluir que a idade dos repositórios não influencia na coesão.

4.3 Releases(atividade)

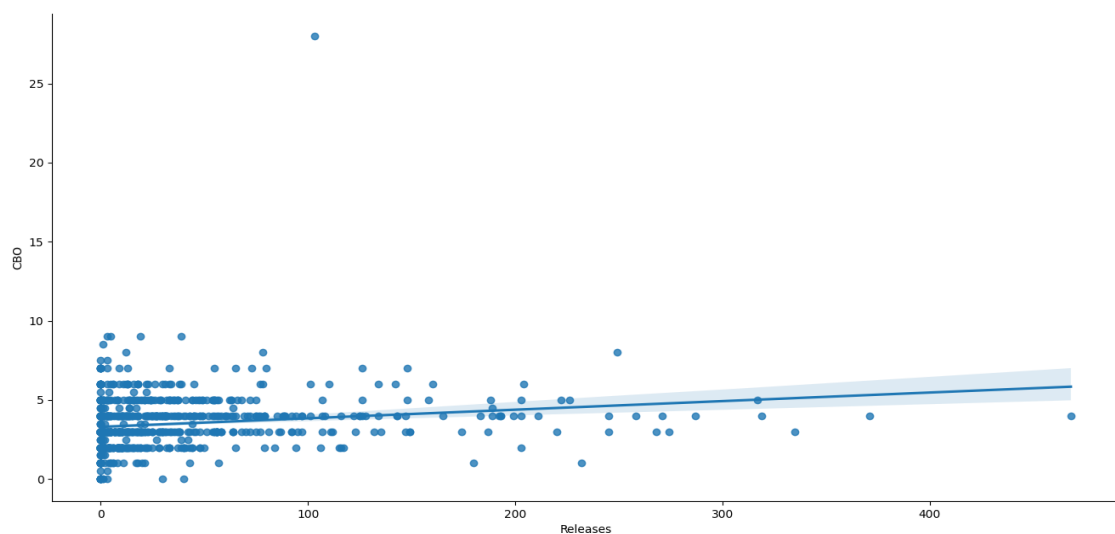


Figura 8. Número de releases dos repositórios correlacionadas com a métrica CBO.

Na Figura 8 pode-se observar uma fraca correlação entre o número de releases de um repositório e o acoplamento entre os objetos (CBO).

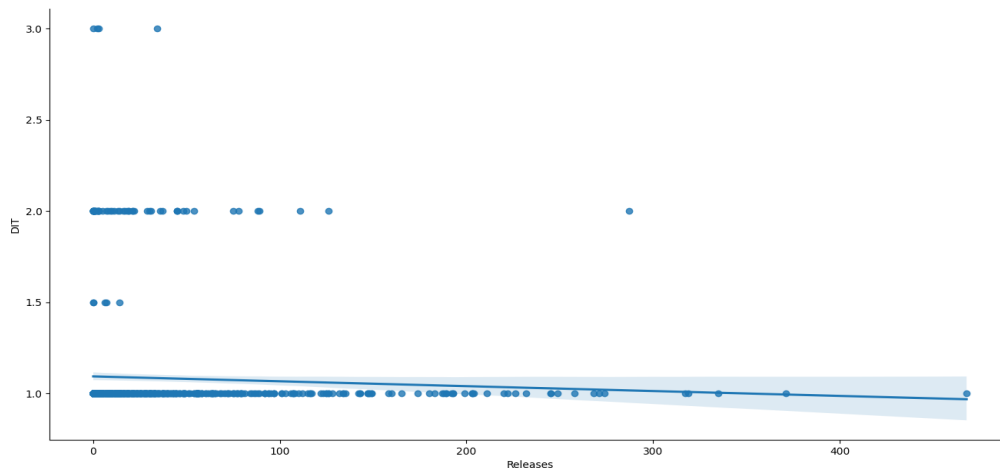


Figura 9. Número de releases dos repositórios correlacionada com a métrica DIT.

Na Figura 9 é possível observar a fraca correlação entre o número de releases de um repositório e a métrica DIT.

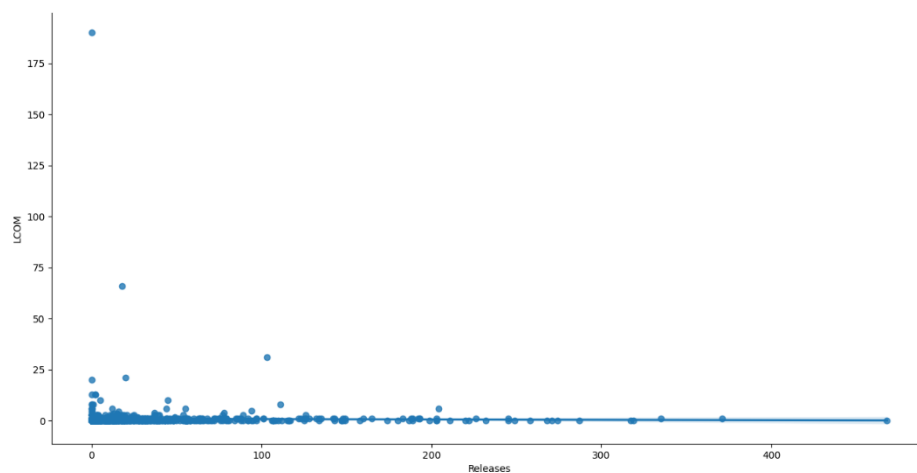


Figura 10. Número de releases dos repositórios correlacionada com a métrica LCOM.

Na Figura 10, é possível observar a fraca correlação entre o número de releases de um repositório com a falta de coesão. Com grande parte dos valores próximos de zero, é possível concluir que a quantidade de releases em repositórios não influencia na falta de coesão.

4.4 Tamanho

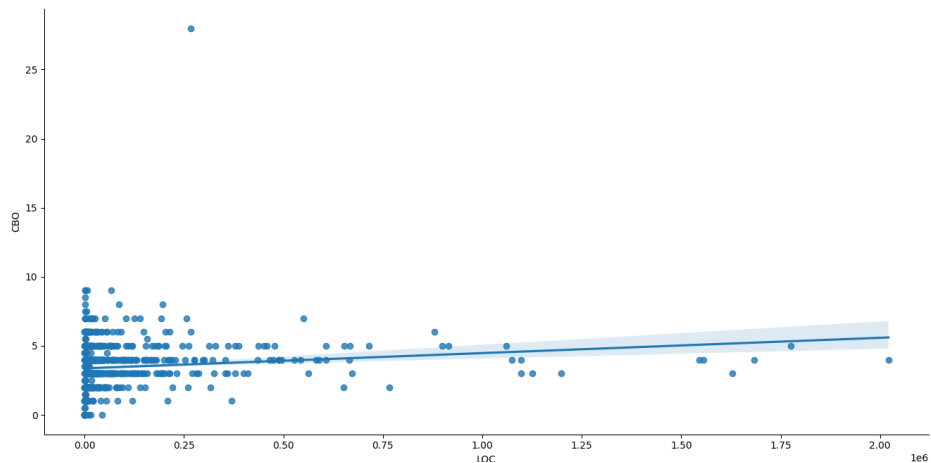


Figura 11. Tamanho dos repositórios relacionada com a métrica CBO

Na Figura 11 é possível observar uma fraca correlação entre o tamanho dos repositórios e o acoplamento entre os objetos (CBO).

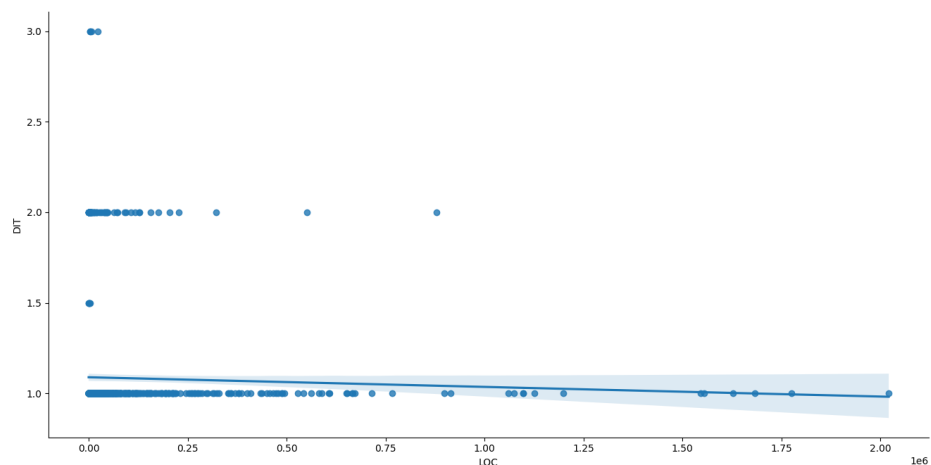


Figura 12. Tamanho dos repositórios relacionada com a métrica DIT.

Na Figura 12, pode-se observar a fraca correlação entre o tamanho do repositório e a métrica DIT. Com a maior parte dos valores próximos de 0, é possível concluir que o tamanho dos repositórios não influencia no número de níveis que uma classe herda.

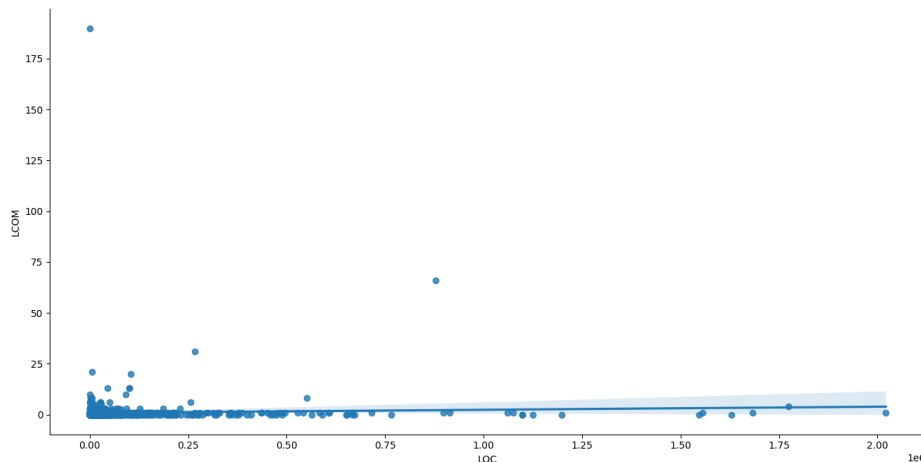


Figura 13. Tamanho dos repositórios correlacionado com a métrica LCOM.

A partir da Figura 13 pode-se observar a fraca correlação entre o tamanho dos repositórios e a falta de coesão. Tendo a média de valores próximo de zero, podemos concluir que o tamanho dos repositórios não influencia na falta de coesão.

5. Discussão

Nesta seção, discutiremos os resultados obtidos com base nas métricas definidas e analisadas. O objetivo é explorar os insights proporcionados pelos dados coletados e avaliar as relações entre os diferentes fatores investigados.

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

A questão tem por objetivo investigar a relação da popularidade de um repositório e suas características de qualidade. Para avaliação desses dados como métricas de qualidade temos DIT, LCOM, CBO, e como parâmetro a popularidade dos repositórios, medida pela quantidade de estrelas.

A hipótese inicial levantada era de que repositórios populares possuem características de qualidade elevadas. Os resultados obtidos não confirmam a hipótese inicial, uma vez que trazem uma baixa correlação entre estes dados com destaque para o LCOM no qual a maior parte dos valores foram muito próximos de zero.

Portanto, não é possível estabelecer uma relação entre a popularidade de repositórios e suas características de qualidade.

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

O objetivo da questão é investigar qual o impacto da maturidade dos repositórios em suas características de qualidade. Para isso, foi utilizado como parâmetro a idade do repositório em anos e, como parâmetros de qualidade, DIT, CBO, LCOM.

A hipótese inicial levantada era de que sistemas mais maduros tem suas características de qualidade mais desenvolvidas. Nas análises realizadas foi possível recusar a hipótese inicial, uma vez que os resultados mostram uma baixa correlação entre os dados. Com destaque para os valores de DIT e LCOM mostrados na Figura 6 e Figura 7 que apresenta grande concentração dos valores próximos de zero, indicando que não se pode apontar uma relação da idade dos repositórios com as métricas de qualidade.

Portanto, é possível concluir que a maturidade do repositório não influencia diretamente nas características de qualidade.

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

O propósito desta questão é investigar a relação entre a atividade de um repositório e suas características de qualidade. Para a definição do parâmetro de atividade foi utilizado o número de releases de um repositório e como parâmetro das características de qualidade DIT, CBO e LCOM.

A hipótese inicial levantada é de que um maior número de releases expressa maior número nas métricas de qualidade. Os resultados obtidos não confirmam a hipótese inicial, uma vez que não foi identificada uma relação significativa entre a atividade do repositório e as métricas utilizadas para medir a qualidade.

Assim, é possível concluir que a atividade do repositório, medida pelo número de releases não é um indicador significativo de correlação com as métricas de qualidade analisadas.

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

A questão busca relacionar o tamanho de um repositório com as suas características de qualidade. Para isso foi utilizado como parâmetro para o tamanho a medição de LOC e para as métricas de qualidade temos como parâmetro DIT, LCOM, CBO.

A hipótese inicial levantada é de que repositórios de tamanhos moderados tem características de qualidade maiores. Os resultados obtidos corroboram parcialmente com as hipóteses iniciais levantadas, tendo um aumento do CBO e do LCOM conforme o aumento do tamanho, apresentando então, declínio da qualidade conforme aumento do tamanho.

