

Motivation

- Quantifying collaboration characteristics of an Open-Source Software (OSS) team and investigating its correlation with the quality of the software artifact produced by the team can lead to predictive models for early detection of potentially poor-quality code in software development projects

Introduction

- To understand how OSS team collaboration relates to the quality of the software produced, we applied text and social analytics to a dataset of 168 OSS projects
- We propose metrics to quantify collaboration characteristics of a team using online communications between team members and some metadata, and measure the quality of software artifacts using SonarQube, a static code analysis tool

Collaboration characteristics

We propose five metrics to measure collaboration characteristics of a team.

- Contribution disparity** to measure disparity of individual's contributions from team average
- Response timeliness** to measures promptness of response on online discussion forums
- Sentiment disparity** to measure disparity of sentiments in online communications between team members
- Team competence as a measure of trust** to measure trust among team members based on past commitment fulfilment; and
- Interaction disparity** to measure disparity of individual's participation in online discussions from team average

Disparities between team members are computed using Gini-coefficient.

Code Quality is measured using *SonarQube* that verifies the code's compliance against specified coding rules, whose violation indicates the software potentially requires maintenance effort.

Results

- Team Competence shows a statistically significant correlation with the code quality
- Interaction disparity and response timeliness show a weak correlation but a medium effect on code quality
- Contribution disparity and sentiment disparity have a negligible correlation and a low effect on code quality

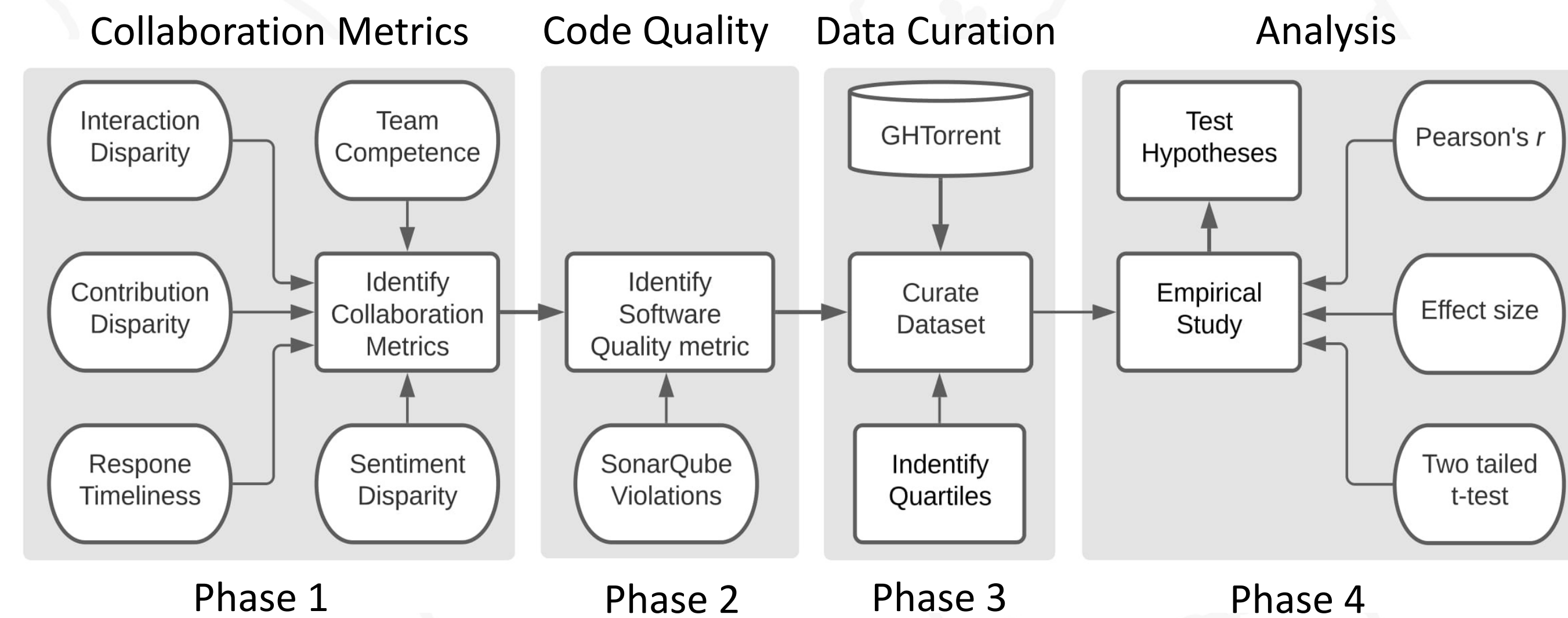


Fig 1. Methodology

Collab Metric	\bar{Q}_1	\bar{Q}_4	$\sigma(Q_1)$	$\sigma(Q_4)$	Pearson's r	Cohen's d	Ttest p -value
Contribution Disp.	164	332	491	607	0.136	0.305	0.166
Response Time	132	307	209	569	0.103	0.416	0.08
Sentiment Disp.	170	371	333	763	0.067	0.341	0.124
Team Competence	410	93	698	213	-0.229	0.611	0.007
Interaction Disp.	117	417	250	736	0.212	0.549	0.017

Table 1. Results comparing code quality produced by teams with low (Q1) and high (Q4) collaboration characteristic

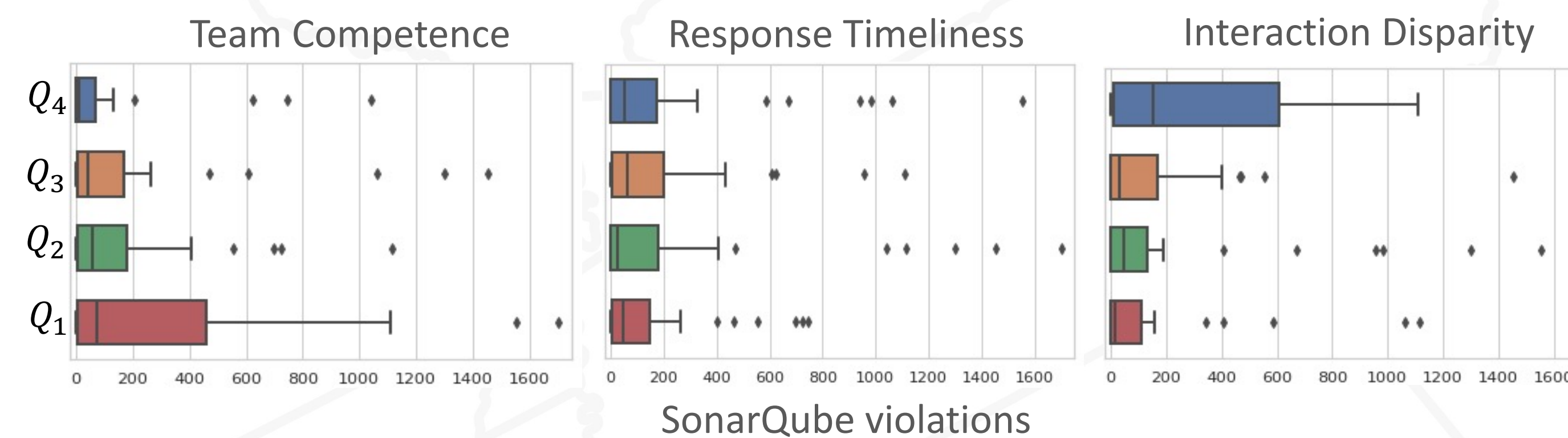


Fig 2. Box plots showing collaboration characteristics against SonarQube violations for teams in each quartile

Methodology

- We propose null hypotheses (collaboration characteristics of an OSS team do not influence the quality of the code produced by the team) and run t-tests and effect size analysis (separately for each proposed metric) to investigate if the correlation between our proposed metrics and code quality is statistically significant
- In order to compare these effects, we split the repositories into quartiles based on computed collaboration metrics and compare the Low (Q1) and High (Q4) quartiles based on code quality

Dataset

- We used the GHTorrent project to identify repositories relevant to this study based on exclusion and inclusion criteria and used Git Developer API to get the data
- Our dataset contains 168 OSS repositories with communication information between the team members (online discussion forums), code changes (commits and commit messages), and the measure of code quality of the software artifact

Conclusion and Future Work

- Our findings highlight the significance of effective communication in team collaboration and identifies some key attributes that can potentially be used in predictive models for early bug identification in software projects
- Particularly, the promptness in responding to online interactions and fixing issues, and efficient communication among team members contribute positively towards code quality

