

**MASARYK  
UNIVERSITY**

FACULTY OF INFORMATICS

# **Source Code Quality impact on Pull Requests acceptance**

Master's Thesis

**ONDŘEJ KUHEJDA**

Advisor: Assistant professor Bruno Rossi

Department of Computer Systems and Communications

Brno, Spring 2022

MUNI  
FI

## **Declaration**

Hereby I declare that this paper is my original authorial work, which I have worked out on my own. All sources, references, and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

Ondřej Kuhejda

**Advisor:** Assistant professor Bruno Rossi

## **Acknowledgements**

I would like to thank my supervisor, Bruno Rossi, for his guidance throughout the whole process.

## **Abstract**

TODO

## **Keywords**

code quality, pull request, static code analysis

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem statement . . . . .	1
<b>2</b>	<b>TODO Pull-based development</b>	<b>2</b>
<b>3</b>	<b>TODO Code quality</b>	<b>3</b>
<b>4</b>	<b>Pull request acceptance</b>	<b>4</b>
4.1	Repository level . . . . .	4
4.2	Submitter level . . . . .	5
4.3	Pull request level . . . . .	5
4.4	<b>TODO</b> Code quality . . . . .	6
4.5	<b>TODO</b> Unsorted . . . . .	7
<b>5</b>	<b>TODO Data sets</b>	<b>8</b>
<b>6</b>	<b>Evaluation</b>	<b>9</b>
<b>7</b>	<b>Threads to validity</b>	<b>10</b>
<b>8</b>	<b>Conclusion</b>	<b>11</b>
8.1	Future work . . . . .	11
	<b>Bibliography</b>	<b>12</b>

# 1 Introduction

## 1.1 Problem statement

- Is there some correlation between the source code quality and the pull request acceptance?
- Which issues/code smells affect the PR acceptance most?
- Is code quality impact higher in projects that are using some particular programming language?
- Is there a better code quality in projects which are using linters as part of their CI?
- Is there a better code quality in projects which are using linters as part of their build process?
- Does code smells lead to a longer discussion about PR?

## 2 TODO Pull-based development

- Wait for It: Determinants of Pull Request Evaluation Latency on GitHub [1]
  - CI and latency



### **3 TODO Code quality**

Trautsch et al. [2] analyzed several open-source projects in regards to usage of static analysis tools. They found out that incorporating a static analysis tool in a build process reduces the defect density.

## 4 Pull request acceptance

Pull request acceptance is a problem that has been studied multiple times. Several surveys were performed in order to understand why pull requests are being rejected.

Gousios et al. [3] surveyed hundreds of integrators to find out their reasons behind the PR rejection. Code quality was stated as the main reason by most of the integrators; code style was in the second place. Factors that integrators examine the most when evaluating the code quality are style conformance and test coverage.

Kononenko et al. [4] performed a study of an open-source project called *Shopify*; they manually analyzed PR's and also surveyed *Shopify* developers. They found out that developers associate the quality of PR with the quality of its description and with the revertability and complexity of the PR.

The reasons why contributors abandon their PRs were also studied [5]. The reason number one was the “Lack of answers from integrators.”; moreover, the “Lack of time” and the “Pull request is obsolete” was also often stated as the main reason.

Even though the different open-source communities solve the problem of pull request acceptance in a different manner, three main governance styles can be identified — protective, equitable, lenient. Protective governance style values trust in the contributor-maintainer relationship. The equitable governance style tries to be unbiased towards the contributors, and the lenient style prioritizes the growth and openness of the community [6]. Each style focuses on different aspects of PR. Tsay et al. [7] identified the following levels of social and technical factors that influence the acceptance of the PR — *repository level*, *submitter level*, and the *pull request level*.

### 4.1 Repository level

The *repository level* is interested in the aspects of the repository itself, such as the repository age, number of collaborators, or number of stars on the GitHub.

For instance, the programming language used in the project also influences the acceptance of the PRs. Pull requests containing Java,

JavaScript, or C++ code have a smaller chance to be accepted than PRs containing the code written in Go or Scala [8].

Furthermore, older projects and projects with a large team have a significantly lower acceptance rate [7].

The popularity of the project also influences the acceptance rate — projects with more stars have more rejected PRs [7].

## 4.2 Submitter level

The *submitter level* is concerned about the submitter's status in the general community and his status in the project itself. There are several parameters that can be considered when evaluating the submitter's status.

PRs of submitters with higher social connection to the project have a higher probability of being accepted [7].

Submitter status in the general community plays an important role in PR acceptance. If the submitter is also a project collaborator, the likelihood that the PR will be accepted increases by 63.3% [7].

Moreover, users that contributed to a larger number of projects have a higher chance that their PR will be accepted [9].

The gender of the submitter is another factor that plays a role in PR acceptance. A study showed that woman's PR are accepted more often, but only when they are not identifiable as a woman [10].

Personality traits also influence PR acceptance. The *IBM Watson Personality Insights* were used to obtain the personality traits of the PR submitters by analyzing the user's comments. These traits were then used to study PR acceptance. It has been shown that conscientiousness, neuroticism, and extroversion are traits that have positive effects on PR acceptance. The chance that PR will be accepted is also higher when the submitter and closer have different personalities [11].

## 4.3 Pull request level

The *pull request level* is interested in the data that are connected to the PR itself. For instance, on the *PR level*, one can study if there is a correlation between PR acceptance and the number of GitHub

comments in the PR. Another parameter that can be used is “Number of Files Changed” or “Number of Commits”.

One of the factors that negatively influence the acceptance rate is the already mentioned number of commits in the pull request. The high number of commits decreases the probability of acceptance. On the other hand, PR’s with only one commit are exceptions — they have a smaller chance to be accepted than pull requests which contain two commits [9].

Another observation is that more discussed PR’s has a smaller chance to be accepted [7]. Another study did not find a large difference between accepted and rejected PR’s based on the number of comments but found that discussions in rejected PR’s have a longer duration [12].

Proper testing is the crucial part of every project, and therefore it also influences the pull request acceptance. PR’s including more tests have a higher chance to be accepted, and an increasing number of changed lines decreases the likelihood of PR acceptance [7].

Testing plays a significant role in discovering bugs and therefore leads to higher code quality. On the other hand, many test cases do not have to mean that code has a high quality. The code quality is an essential factor on the *pull request level*, therefore, is this study’s main interest. Works that are also interested in the code quality and the pull request acceptance are examined in the following chapter.

Another factor that is closely tied to code quality is the code style. This factor has a small (but not negligible) negative effect on acceptance. This means that PRs with larger code style inconsistency (with the codebase) have a smaller chance of being accepted [13].

#### 4.4 TODO Code quality

Although most integrators view code quality as the most important factor regarding PR acceptance, to the best of my knowledge, only one study was performed to discover whether there is a connection between the PR’s acceptance and its quality.

- Does code quality affect pull request acceptance? [14]

### 4.5 TODO Unsorted

- study “Influence of Social and Technical Factors” [7] was replicated [11]
- Replication Can Improve Prior Results: A GitHub Study of Pull Request Acceptance [15]
  - contains interesting table with factors that influences acceptance
- Pull Request Decision Explained: An Empirical Overview [16]
  - also contains interesting table with factors that influences acceptance
- An Exploratory Study of the Pull-Based Software Development Model [17]
- Which Pull Requests Get Accepted and Why? A study of popular NPM Packages [18]
- Rejection Factors of Pull Requests Filed by Core Team Developers in Software Projects with High Acceptance Rates [19]
- Pull Request Prioritization Algorithm based on Acceptance and Response Probability [20]

## 5 TODO Data sets

- <https://dl.acm.org/doi/abs/10.1145/2597073.2597122>
- <https://dl.acm.org/doi/abs/10.1145/3379597.3387489>
- <https://zenodo.org/record/3858046>
- <https://github.com/XLipcak/rev-rec>
- <https://ghtorrent.org/>
- Kalliamvakou et al. noted that data about PR's mined from GitHub are not always reliable, because PR can be also merged using several different approaches.
  - <https://dl.acm.org/doi/10.1145/2597073.2597074>
  - [17]

## 6 Evaluation

## **7 Threads to validity**



## **8 Conclusion**

### **8.1 Future work**

## Bibliography

1. YU, Y.; WANG, H.; FILKOV, V.; DEVANBU, P.; VASILESCU, B. Wait for It: Determinants of Pull Request Evaluation Latency on GitHub. In: *Working Conference on Mining Software Repositories*. IEEE, 2015, pp. 367–371. Available from DOI: 10.1109/MSR.2015.42.
2. TRAUTSCH, A.; HERBOLD, S.; GRABOWSKI, J. A longitudinal study of static analysis warning evolution and the effects of PMD on software quality in Apache open source projects. *Empirical Software Engineering*. 2020, vol. 25, pp. 5137–5192. Available from DOI: 10.1007/s10664-020-09880-1.
3. GOUSIOS, G.; ZAIDMAN, A.; STOREY, M.; DEURSEN, A. Work Practices and Challenges in Pull-Based Development: The Integrator’s Perspective. In: *International Conference on Software Engineering*. IEEE, 2015, vol. 1, pp. 358–368. Available from DOI: 10.1109/ICSE.2015.55.
4. KONONENKO, O.; ROSE, T.; BAYSAL, O.; GODFREY, M.; THEISEN, D.; WATER, B. Studying Pull Request Merges: A Case Study of Shopify’s Active Merchant. In: *International Conference on Software Engineering*. ACM, 2018, pp. 124–133. Available from DOI: 10.1145/3183519.3183542.
5. LI, Z.; YU, Y.; WANG, T.; YIN, Gang; LI, Shanshan; WANG, Huaimin. Are You Still Working on This An Empirical Study on Pull Request Abandonment. *Transactions on Software Engineering*. 2021, pp. 1–1. Available from DOI: 10.1109/TSE.2021.3053403.
6. ALAMI, A.; COHN, L.; WĄSOWSKI, A. How Do FOSS Communities Decide to Accept Pull Requests? In: *Proceedings of the Evaluation and Assessment in Software Engineering*. ACM, 2020, pp. 220–229. Available from DOI: 10.1145/3383219.3383242.
7. TSAY, J.; DABBISH, L.; HERBSLEB, J. Influence of Social and Technical Factors for Evaluating Contribution in GitHub. In: *International Conference on Software Engineering*. ACM, 2014, pp. 356–366. Available from DOI: 10.1145/2568225.2568315.

8. SOARES, D.; DE LIMA, M.; MURTA, L.; PLASTINO, A. Acceptance Factors of Pull Requests in Open-Source Projects. In: *Symposium on Applied Computing*. ACM, 2015, pp. 1541–1546. Available from DOI: 10.1145/2695664.2695856.
9. DEY, T.; MOCKUS, A. Effect of Technical and Social Factors on Pull Request Quality for the NPM Ecosystem. In: *International Symposium on Empirical Software Engineering and Measurement*. ACM, 2020. Available from DOI: 10.1145/3382494.3410685.
10. JOSH, J.; KOFINK, A.; MIDDLETON, J.; RAINEAR, C.; MURPHY-HILL, E.; PARNIN, C.; STALLINGS, J. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science*. 2017, vol. 3. Available from DOI: 10.7717/peerj-cs.111.
11. IYER, R.; YUN, A.; NAGAPPAN, M.; HOEY, J. Effects of Personality Traits on Pull Request Acceptance. *Transactions on Software Engineering*. 2019. Available from DOI: 10.1109/TSE.2019.2960357.
12. GOLZADEH, M.; DECAN, A.; MENS, T. On the Effect of Discussions on Pull Request Decisions. In: *Belgium-Netherlands Software Evolution Workshop*. CEUR Workshop Proceedings, 2019. Available also from: <http://ceur-ws.org/Vol-2605/16.pdf>.
13. ZOU, W.; XUAN, J.; XIE, X.; CHEN, Z.; XU, B. How does code style inconsistency affect pull request integration? an exploratory study on 117 github projects. *Empirical Software Engineering*. 2019, vol. 24, pp. 3871–3903. Available from DOI: 10.1007/s10664-019-09720-x.
14. LENARDUZZI, V.; NIKKOLA, V.; SAARIMÄKI, N.; TAIBI, D. Does code quality affect pull request acceptance? An empirical study. *Journal of Systems and Software*. 2021, vol. 171, pp. 110806. Available from DOI: 10.1016/j.jss.2020.110806.
15. CHEN, D.; STOLEE, K.; MENZIES, T. Replication Can Improve Prior Results: A GitHub Study of Pull Request Acceptance. In: *International Conference on Program Comprehension*. IEEE, 2019, pp. 179–190. Available from DOI: 10.1109/ICPC.2019.00037.

## BIBLIOGRAPHY

---

16. ZHANG, X.; YU, Y.; GOUSIOS, G.; RASTOGI, A. Pull Request Decision Explained: An Empirical Overview. *Computing Research Repository*. 2021. Available from arXiv: 2105.13970.
17. GOUSIOS, G.; PINZGER, M.; DEURSEN, A. An Exploratory Study of the Pull-Based Software Development Model. In: *International Conference on Software Engineering*. ACM, 2014, pp. 345–355. Available from DOI: 10.1145/2568225.2568260.
18. DEY, T.; MOCKUS, A. Which Pull Requests Get Accepted and Why? A study of popular NPM Packages. *Computing Research Repository*. 2020. Available from arXiv: 2003.01153.
19. SOARES, D.; DE LIMA, M.; MURTA, L.; PLASTINO, A. Rejection Factors of Pull Requests Filed by Core Team Developers in Software Projects with High Acceptance Rates. In: *International Conference on Machine Learning and Applications*. IEEE, 2015, pp. 960–965. Available from DOI: 10.1109/ICMLA.2015.41.
20. AZEEM, I.; PENG, Q.; WANG, Q. Pull Request Prioritization Algorithm based on Acceptance and Response Probability. In: *International Conference on Software Quality*. IEEE, 2020, pp. 231–242. Available from DOI: 10.1109/QRS51102.2020.00041.