Examining Changes in Good First Issue Practices and Newcomer Pull Request Characteristics in Popular OSS Projects

Abstract—**Open-source software (OSS) projects rely on effective newcomer onboarding to sustain their communities. Many projects use "good first issue" (GFI) labels to highlight beginner-friendly tasks. As development practices continue to evolve, understanding how these onboarding mechanisms change over time is important for both maintainers and researchers. This study analyzes 406,826 issues and 1,117 PRs addressing GFIs across 37 popular GitHub repositories (30 of which use GFI labels) over a four-year period from July 2021 to June 2025. Using time-series analysis with Mann-Kendall trend tests, we find that while the proportion of issues with GFI labels remained stable during the first three years, it declined sharply in the fourth year (Kendall's $\tau = -0.44$, $p < 0.001$), with substantial variation across projects. This heterogeneity was not explained by repository age or primary programming language. Notably, newcomer engagement with GFI issues remains stable at approximately 27% throughout the period, suggesting that GFI labels maintain consistent attractiveness for newcomers. A label-based task type analysis reveals that bug-fix GFIs have the highest merge rate (68.7%), while the "Other" category shows a steep decline from 57.1% to 28.6%, partly driven by project-specific labeling practices. Initial PR characteristics such as description length and code size show no significant association with merge outcomes, suggesting that GFIs are appropriately scoped for newcomers. Our findings suggest that GFI trends are driven by project-specific strategic decisions rather than external factors. These results highlight that GFI trends are shaped by project-level decisions rather than ecosystem-wide factors, offering actionable insights for maintainers and newcomers.**

Index Terms—**Open source Software, Newcomer onboarding, Good first issue, GitHub**

I. Introduction

The sustained development of open-source software (OSS) projects depends critically on the influx and retention of new contributors [?], [?]. However, newcomers often face significant technical and social barriers when attempting to make their first contribution, including difficulty finding suitable tasks, understanding complex codebases, and navigating unfamiliar development processes [?], [?].

To address these challenges, many OSS projects have adopted the practice of labeling certain issues as "good first issue" (GFI) to identify tasks suitable for beginners. Prior research has examined the usage patterns of GFI labels and proposed automated recommendation methods [?], [?]. However, Tan et al. [?] showed that many GFIs are not resolved by newcomers, indicating that challenges remain regarding the effectiveness of GFI labeling practices.

In recent years, the software development landscape has undergone substantial changes, including the widespread adoption of remote work and the emergence of new development tools. Understanding how GFI practices and newcomer behavior have evolved over time is important for maintaining healthy OSS communities.

However, **longitudinal trends in GFI practices and the behavior of newcomers who engage with GFI issues remain unclear**. While prior work has examined GFI effectiveness and recommendation methods at specific points in time, we lack understanding of how GFI practices have evolved over recent years and how the characteristics and success rates of GFI-related contributions have changed.

In this study, we analyze four years (July 2021 to June 2025) of GFI practices across 37 popular OSS projects on GitHub. We address the following research questions:

RQ1: How have GFI practices and newcomer engagement changed over the four-year period?

We examine the trends in GFI issue proportion and newcomer engagement rates, investigating how these patterns have evolved over time.

RQ2: How do task type labels of GFI issues relate to newcomer PR merge outcomes, and what factors are associated with merge success?

We classify GFI issues into task types (Bug, Feature, Documentation, Other) based on their labels and analyze how merge rates differ by task type and over time. We also examine PR-level factors associated with merge success.

Through an analysis of 406,826 issues (identifying 3,300 GFI-labeled issues) and 1,117 GFI PRs, we find that the proportion of GFI issues remained stable for the first three years but declined sharply in the fourth year ($\tau = -0.44$, $p < 0.001$), with substantial variation across projects. Notably, newcomer engagement with GFI issues remains stable at approximately 27% throughout the four-year period, though 73% of GFIs remain unaddressed by newcomers. Our findings provide empirical evidence on the evolution of GFI practices and offer practical insights for OSS communities seeking to improve newcomer onboarding.

II. Related Work

Newcomer Onboarding in OSS. Steinmacher et al. [?], [?] systematically classified the barriers that newcomers face when joining OSS projects, highlighting the importance of social as well as technical barriers. Subramanian et al. [?] analyzed the characteristics of newcomers' first contributions, revealing that approximately half of them address bug fixes.

Turzo et al. [?] evaluated the effectiveness of onboarding recommendations and showed that effective strategies vary by project. Steinmacher et al. [?] proposed FLOSScoach, a portal for newcomers, and demonstrated its effectiveness in reducing orientation barriers.

Good First Issues and Task Recommendation. Tan et al. [?] were the first to systematically analyze the use of the Good First Issue (GFI) label on GitHub, showing that many GFIs are not resolved by newcomers. Xiao et al. [?] and Huang et al. [?] proposed methods for automatically recommending GFIs using machine learning. Santos et al. [?] revealed a mismatch in task selection strategies between newcomers and existing developers. Xiao et al. [?] proposed a personalized task recommendation method based on contributor background.

Mentoring and Community Support. Steinmacher et al. [?] and Balali et al. [?] identified the challenges and strategies of task recommendation from the perspective of OSS mentors. Cao et al. [?] showed that simply labeling GFIs is insufficient and that direct support from mentors is crucial for newcomer success. Guizani et al. [?] proposed a maintainer dashboard to support the attraction of new contributors.

Building on this prior work, our study empirically analyzes longitudinal trends in GFI usage and newcomer engagement over a four-year period. While prior studies examined GFI effectiveness at specific points in time, our work contributes by analyzing temporal trends using time-series methods.

III. METHODOLOGY

Figure **??** illustrates the overview of our research methodology. We selected our target projects from the top 50 most-starred repositories on GitHub, collected issues and newcomer pull requests, and analyzed GFI label trends and newcomer engagement (RQ1) and PR characteristics with merge success factors (RQ2).

A. Repository Selection

To focus on projects with active contributions, we only selected repositories with 50 or more pull requests per year. Furthermore, we excluded non-software projects (e.g., tutorial collections, learning resources) and projects that have disabled GitHub Issues (e.g., django/django), resulting in a final set of 37 software repositories for analysis, of which 30 used GFI labels.

The selected repositories are diverse, with the distribution of primary programming languages being TypeScript (9 projects, 24.3%), C++ (5, 13.5%), JavaScript (5, 13.5%), Python (5, 13.5%), Rust (3, 8.1%), and others (10, 27.0%).

B. Data Collection

1) Issue Data Collection (RQ1): To answer RQ1, we used the GitHub GraphQL API to collect all GFI-labeled issues from July 2021 to June 2025. To identify GFIs, we followed the method adopted by Turzo et al. [?], combining the label list presented by Tan et al. [?] with newcomer contribution guidelines [?]. For each issue, we recorded its creation date, label information, and closed state.

To calculate GFI label usage rates, we collected 406,826 total issues and identified 3,300 issues with GFI labels. Additionally, to analyze newcomer engagement with GFI tasks, we collected pull requests addressing these GFI issues.

2) Pull Request Data Collection (RQ2): To answer RQ2, we collected pull requests that address issues with GFIs (hereafter referred to as GFI PRs) using the GitHub GraphQL API. Newcomers were identified on a per-repository basis: we extracted the first-ever pull request submitted by each user to a given repository between July 2021 and June 2025. For each pull request, we retrieved the PR number, title, body, creation date, merge date, state (MERGED, CLOSED, OPEN), lines added, lines deleted, number of changed files, commit count, review comment count, and label information. For the description, we measured the length of substantive user-written content after removing HTML comments from PR templates. To ensure data quality, we filtered out bots and deleted accounts using the author_type field from the GitHub API, retaining only those where author_type was 'User'. All data was collected via the GitHub API in November 2025, approximately five months after the end of the study period. This observation buffer exceeds the 95th percentile of time-to-merge among merged PRs (approximately 80 days), ensuring that PRs created near the end of the analysis window had sufficient time to be reviewed and resolved. For merge rate calculations, we treated all non-MERGED PRs (including 68 still-open PRs) as not merged. For metrics with highly skewed distributions, we applied a log transformation to insertions and deletions.

For time-series comparison, we divided the four-year period into 12-month analysis years: Y1 (Jul 2021–Jun 2022), Y2 (Jul 2022–Jun 2023), Y3 (Jul 2023–Jun 2024), and Y4 (Jul 2024–Jun 2025). We classified each PR into a task type based on the labels of its referenced GFI issue: Bug (label contains "bug"), Feature ("feature" or "enhancement"), Documentation ("doc"), and Other (none of the above). Word-boundary matching for "bug" prevents false positives from area labels such as "debug." Of the 1,117 PRs, 1,070 (95.8%) matched exactly one task type and were classified automatically. The remaining 47 PRs (4.2%) matched multiple types; the first author manually classified each case by distinguishing type labels (e.g., "type: bug", "C-enhancement", "documentation") from area/module labels (e.g., "addon: docs", "A-docs", "module: docs"). For example, a PR with "type: bug" and "docs" (area) was classified as Bug, while a PR with both "enhancement" and "documentation" as type labels was classified as Documentation. The full list of 47 overlapping cases with their manual classifications is included in the replication package.

The final dataset consists of 43,906 first pull requests from newcomers and 1,117 GFI PRs.

IV. RESULTS

A. RQ1: How have GFI practices and newcomer engagement changed over time?

1) GFI Ratio Trend: We analyzed the monthly GFI ratio over the four-year period (July 2021 to June 2025). A

Mann-Kendall trend test indicated a statistically significant decreasing trend ($\tau = -0.44$, $p < 0.001$). However, as shown in Figure **??**(a), this decline was not gradual: the yearly average GFI ratio remained stable from Y1 (0.92%) through Y3 (0.88%), then dropped sharply in Y4 (0.57%).

2) Repository Heterogeneity in GFI Trends: Of the 37 repositories, 7 (18.9%) never used GFI labels during the analysis period. We conducted Mann-Kendall trend tests on the remaining 30 repositories. These repositories showed substantial heterogeneity in GFI usage trends: 10 (33.3%) showed a decreasing trend, 17 (56.7%) showed no significant trend, and 3 (10.0%) showed an increasing trend. This variation could not be explained by repository age (Spearman's $\rho = -0.105$, $p = 0.588$) or primary programming language.

Comparing characteristics across the three trend groups, no statistically significant differences were found in star count, repository age, total issue count, or GFI count (all $p > 0.2$). These results suggest that GFI usage trends cannot be explained by objective characteristics such as project size, maturity, or primary language, but rather depend heavily on project-specific strategic decisions.

3) Newcomer Engagement with GFIs: As shown in Figure **??**(b), the proportion of GFI issues addressed by newcomers remained stable throughout the period. The overall engagement rate was 27.0% (891 out of 3,300 GFI issues), with per-year rates between 25% and 29%. A Mann-Kendall trend test confirmed no significant trend ($\tau = 0.06$, $p = 0.52$). Despite the sharp decline in GFI ratio observed in Y4 (Figure **??**(a)), the stable engagement rate (Figure **??**(b)) suggests that GFI labels maintain a consistent level of attractiveness for newcomers regardless of their prevalence.

B. RQ2: How do task type labels relate to merge outcomes, and what factors are associated with merge success?

1) PR Metrics Trends and Task Type Analysis: We analyzed 1,117 GFI-labeled PRs (after excluding bots). The overall merge rate was 53.0%. Table **??** summarizes the time-series trends for key metrics. The merge rate showed a statistically significant decreasing trend ($\tau = -0.35$, $p < 0.001$), while description length showed an increasing trend ($\tau = 0.35$, $p < 0.001$).

We analyzed merge rates by task type over four analysis years (Table **??**). Bug-fix tasks had the highest overall merge rate (68.7%), peaking at 83.5% in Y2 before declining to 45.9% in Y4. Feature tasks remained stable at approximately 54% with no significant trend. The "Other" category, which includes PRs whose GFI issues lack standard task-type labels (e.g., issues labeled only with module or area tags), showed the steepest decline from 57.1% (Y1) to 28.6% (Y4). This decline was partly driven by a single project (PyTorch) that contributed 188 PRs with 0% merge rate due to its module-based labeling system.

2) Factors Associated with Merge Success: Table **??** compares merged and unmerged GFI PRs. Among initial PR characteristics—code size, number of changed files, and description length—none showed a statistically significant association with merge outcomes. This suggests that GFIs are appropriately scoped for newcomers, and that the scale of the contribution or detail of the description at submission time does not predict success.

Table I. Time-series trend analysis of GFI metrics (Mann-Kendall test)

Metric	Y1	Y4	Kendall τ	Trend
GFI Ratio (%)	0.92	0.57	-0.44***	Decreasing
Newcomer Engagement (%)	25.1	27.7	0.06	No trend
Merge Rate (%)	61.9	42.2	-0.35***	Decreasing
Description Length	306	481	0.35***	Increasing

$^{N}ote: ***$p¡0.001.$MonthlyMann-Kendalltestover48months.Y1/Y4areyearlyaveragesofmonthlyvalues.$

Table II. Merge rate by task type and analysis year

Task Type	Y1	Y2	Y3	Y4	Total	Trend
Bug	64.5%	83.5%	71.9%	45.9%	68.7%	Decr.**
Feature	53.7%	54.8%	53.3%	55.6%	54.4%	None
Docs	68.4%	65.2%	42.4%	47.7%	52.9%	Decr.*
Other	57.1%	46.4%	33.3%	28.6%	40.7%	Decr.*

Note:

Y1=Jul'21–Jun'22, Y2=Jul'22–Jun'23, Y3=Jul'23–Jun'24, Y4=Jul'24–Jun'25.
$*p < 0.05$, $**p < 0.01$, $***p < 0.001$ (Mann-Kendall test).

The two process-level metrics (commit count and review count) showed statistically significant differences. However, these metrics accumulate during the review lifecycle and are subject to confounds. In particular, many popular OSS projects enforce branch protection rules that require at least one approving review before a PR can be merged, creating a mechanical correlation between review count and merge status. Abandoned or closed PRs, by contrast, may never enter the review pipeline. We therefore interpret these process-level differences with caution, as they likely reflect the consequences of the merge process rather than predictive factors.

V. Discussion

A. Interpretation of RQ1 Findings

The GFI ratio remained stable from Y1 (0.92%) through Y3 (0.88%) before declining sharply in Y4 (0.57%). While the Mann-Kendall test indicates a significant trend over the full 48-month period ($\tau = -0.44$, $p < 0.001$), the decline is concentrated in the final year rather than representing a gradual four-year decrease. This trend also varied substantially across repositories: 33.3% showed a decreasing trend, 56.7% showed no significant trend, and 10.0% showed an increasing trend. We found no statistically significant differences between these groups regarding repository age or primary programming language, and the change in GFI usage was uncorrelated with project maturity (Spearman's $\rho = -0.105$, $p = 0.588$).

These results suggest that changes in GFI usage are not uniformly driven by a single external factor, but are strongly dependent on project-specific strategic decisions. Some projects may be reducing their use of GFI labels due to changes in their issue triage process, constraints on maintenance resources, or

Table III. GFI PR metrics by merge status

Metric	Merged	Not Merged	p-value	$\lvert r\rvert$
Initial PR characteristics				
Insertions (log)	3.02	2.89	0.885	–
Deletions (log)	1.10	1.39	0.994	–
Changed Files	2.0	2.0	0.155	–
Description Length	382.5	435.0	0.608	–
Process-level metrics†				
Commits Count	3.0	2.0	<0.001***	0.15
Review Count	2.0	1.0	<0.001***	0.32

Note: $^{***}p < 0.001$. Median values shown. Mann-Whitney U test. $\lvert r\rvert$: rank-biserial effect size. †Process-level metrics accumulate during the review lifecycle and are subject to mechanical confounds (see text).

shifts in their community growth strategy. Conversely, projects actively seeking to attract new contributors may be increasing their use of GFIs.

Notably, the proportion of GFIs addressed by newcomers remained stable at approximately 27% throughout the four-year period (no significant trend, $\tau = 0.06$, $p = 0.52$). Despite the GFI ratio decline in Y4, this stable engagement rate suggests that GFI labels maintain a consistent level of attractiveness for newcomers. However, 73% of GFIs did not receive a newcomer PR, consistent with Tan et al. [**?**]'s finding that many GFIs remain unaddressed by newcomers.

B. Interpretation of RQ2 Findings

The analysis for RQ2 revealed a decreasing trend in the merge rate for newcomer GFI PRs. Label-based task type analysis showed that bug-fix tasks had the highest merge rate (68.7%), while documentation tasks showed a decreasing trend (68.4% in Y1 to 47.7% in Y4). The "Other" category exhibited the steepest decline (57.1% to 28.6%), which is partly explained by project-specific labeling practices. Notably, PyTorch uses a module-based labeling system (e.g., "module: dynamo", "module: inductor") rather than task-type labels, contributing 188 GFI PRs with a 0% merge rate. Excluding PyTorch, the "Other" category merge rate was 64.4% (Y1) to 48.0% (Y4)—still declining but less dramatically. This highlights the importance of accounting for project heterogeneity in labeling practices when analyzing GFI effectiveness.

On the other hand, it is interesting that the description length of newcomer PRs increased significantly. This trend may suggest that project quality standards have risen, or that newcomers have learned to provide more detailed descriptions. However, the fact that the merge rate has decreased despite the increase in description length indicates that detailed descriptions alone are not a decisive factor for merge success.

The most notable finding regarding merge success factors is that initial PR characteristics—code size, number of changed files, and description length—showed no significant association with merge outcomes. This suggests that GFIs are appropriately scoped, making the initial scale of the contribution less relevant to success.

While review count and commit count showed statistically significant differences between merged and not-merged PRs, these are process-level metrics subject to important confounds. Many popular projects enforce branch protection rules requiring approving reviews before merge, which creates a mechanical correlation: merged PRs necessarily pass through the review pipeline, while abandoned PRs may not. Additionally, maintainers may allocate more review effort to PRs they consider viable (reverse causality). We therefore do not interpret review count as an independent predictor of merge success. Rather, our results highlight that *what newcomers bring initially* (code volume, description detail) matters less than *what happens during the review process*. This is consistent with Cao et al. [**?**], who found that mentoring—rather than mere task recommendation—is crucial for newcomer success.

C. Implications for Practice

For Project Maintainers: Our findings indicate that GFI labels continue to attract a stable proportion of newcomers (approximately 27% engagement rate throughout the four-year period). As the overall GFI ratio shows a decreasing trend, actively creating and labeling GFIs can provide a competitive advantage in newcomer acquisition. Notably, bug-fix tasks have the highest merge rate (68.7%), suggesting that maintainers should prioritize labeling well-scoped bug fixes as GFIs for newcomers.

For Newcomers: As the proportion of GFIs is declining, newcomers may need to explore multiple projects. Our analysis shows that some projects continue to actively create GFIs. Given that bug-fix tasks have the highest merge rate, newcomers may increase their success probability by selecting bug fixes for their first contribution. We recommend using filtering mechanisms on GitHub Search and GFI aggregators [**?**] to efficiently find GFIs that match your skill level.

Furthermore, RQ2 findings suggest that initial PR characteristics such as code size or description length do not predict merge outcomes, indicating that GFIs are appropriately scoped. For newcomers, the key implication is that the initial quality of the submission matters less than engaging with the review process: proactively seeking and responding to reviewer feedback is recommended, as prior work has shown that mentoring interactions are crucial for newcomer success [**?**].

For Researchers: This study emphasizes the importance of considering project heterogeneity in OSS onboarding research. Analyses based solely on aggregated statistics may overlook the diverse strategies adopted by individual projects. Additionally, our label-based task type analysis demonstrates that more granular categorization can yield actionable insights. However, researchers should be aware that labeling practices vary significantly across projects—some use task-type labels (bug, feature, documentation), while others use module-based or area-based labels that do not fit standard classification schemes. In our dataset, 4.2% of PRs matched multiple task types, requiring manual classification by distinguishing type labels from area labels. This heterogeneity should be accounted for in cross-project studies.

D. Threats to Validity

Construct Validity: In this study, we used the list of GFI labels presented by Tan et al. [**?**] and Turzo et al. [**?**]. However, some projects may adopt their own label naming conventions, which means we may not have captured all GFIs. Our task type classification (Bug, Feature, Documentation, Other) is based on issue label keywords, with 47 overlapping cases (4.2%) manually classified by the first author. Some projects use alternative labeling schemes (e.g., module-based labels) that do not map cleanly to these categories. Additionally, we defined newcomers as 'individuals submitting their first PR to a repository,' which does not take into account their overall experience on GitHub. Our bot detection relies on GitHub's author_type field, which may not capture all automated contributions (e.g., bots configured as regular users or semi-automated tools used by human developers).

Internal Validity: This study observes time-series trends and does not claim specific causal relationships. The observed changes may be influenced by multiple confounding factors, including overall changes in the OSS ecosystem, strategic decisions by individual projects, and the evolution of development tools. Because our GFI label data is a point-in-time snapshot collected in November 2025, one might suspect that the sharp Y4 decline in GFI ratio reflects labeling lag rather than a genuine trend. However, we confirmed that repositories with reduced GFI counts in Y4 continued to create issues at their usual volume (e.g., hundreds of issues per month) yet assigned zero GFI labels for periods of seven to nine months—well beyond plausible triage delays—indicating a real change in labeling practice rather than a data-collection artifact.

External Validity: This study is limited to popular OSS projects (top-starred) on GitHub. Different trends may be observed in smaller projects or on other platforms (e.g., GitLab, Bitbucket). Furthermore, the analysis is confined to software projects, excluding non-software repositories such as documentation and learning resources.

VI. Conclusion

We analyzed 406,826 issues (identifying 3,300 GFI-labeled issues), 43,906 newcomer pull requests, and 1,117 GFI PRs from 37 popular OSS projects on GitHub to investigate the usage rate of GFI labels, newcomer engagement patterns, and the changing characteristics of GFI PRs over a four-year period from July 2021 to June 2025.

Regarding RQ1, the proportion of GFIs remained stable for the first three years before declining sharply in the fourth year ($\tau = -0.44, p < 0.001$), with this trend varying greatly among repositories and being strongly dependent on project-specific strategic decisions. Meanwhile, newcomer engagement with GFI issues remained stable at approximately 27%, suggesting that GFI labels maintain a consistent level of attractiveness for newcomers.

Regarding RQ2, the merge rate of newcomer GFI PRs showed a decreasing trend. Bug-fix tasks had the highest overall merge rate (68.7%), peaking at 83.5% in Y2 before declining to 45.9% in Y4. Feature tasks remained stable at approximately 54%. The steep decline in the "Other" category (57.1% to 28.6%) was partly attributable to project-specific labeling practices, highlighting the importance of accounting for heterogeneity in cross-project analyses. In the analysis of merge success factors, initial PR characteristics such as code size and description length showed no significant association with merge outcomes, suggesting that GFIs are appropriately scoped for newcomers.

The findings of this study offer practical implications for project maintainers designing strategies for attracting new contributors. Given the decline in GFI prevalence observed in Y4, proactively creating and labeling GFIs can provide a competitive advantage for newcomer acquisition. For newcomers, it is important to explore multiple projects and engage with the review process.

Future research would benefit from (1) a qualitative investigation into the decision-making processes behind why some projects are increasing their use of GFIs while others are decreasing it; (2) identification of the root causes for the decline in GFI PR merge rates (e.g., issue quality, maintenance resources, changes in quality standards); and (3) examination of task-type-specific support strategies for newcomers. Such studies are expected to deepen our understanding of effective newcomer onboarding strategies in sustainable OSS ecosystems.

Data Availability

The replication package for this study is available at https://doi.org/10.5281/zenodo.17638558.

References

[1] I. Steinmacher, T. Conte, M. A. Gerosa, and D. Redmiles, "Social barriers faced by newcomers placing their first contribution in open source software projects," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1379–1392. [Online]. Available: https://doi.org/10.1145/2675133.2675215

[2] D. Sholler, I. Steinmacher, D. Ford, M. Averick, M. Hoye, and G. Wilson, "Ten simple rules for helping newcomers become contributors to open projects," *PLoS computational biology*, vol. 15, no. 9, p. e1007296, 2019.

[3] I. Steinmacher, A. P. Chaves, T. U. Conte, and M. A. Gerosa, "Preliminary empirical identification of barriers faced by newcomers to open source software projects," in *2014 Brazilian Symposium on Software Engineering*, 2014, pp. 51–60.

[4] W. Xiao, H. He, W. Xu, X. Tan, J. Dong, and M. Zhou, "Recommending good first issues in github oss projects," in *Proceedings of the 44th International Conference on Software Engineering*, ser. ICSE '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 1830–1842. [Online]. Available: https://doi.org/10.1145/3510003.3510196

[5] Y. Huang, J. Wang, S. Wang, Z. Liu, D. Wang, and Q. Wang, "Characterizing and predicting good first issues," in *Proceedings of the 15th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, ser. ESEM '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3475716.3475789

[6] X. Tan, M. Zhou, and Z. Sun, "A first look at good first issues on github," in *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ser. ESEC/FSE 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 398–409. [Online]. Available: https://doi.org/10.1145/3368089.3409746

[7] V. N. Subramanian, I. Rehman, M. Nagappan, and R. G. Kula, "Analyzing first contributions on github: What do newcomers do?" *IEEE Software*, vol. 39, no. 1, pp. 93–101, 2022.

[8] A. K. Turzo, S. Sultana, and A. Bosu, "From first patch to long-term contributor: Evaluating onboarding recommendations for oss newcomers," *IEEE Trans. Softw. Eng.*, vol. 51, no. 4, p. 1303–1318, Apr. 2025. [Online]. Available: https://doi.org/10.1109/TSE.2025.3550881

[9] I. Steinmacher, T. U. Conte, C. Treude, and M. A. Gerosa, "Overcoming open source project entry barriers with a portal for newcomers," in *Proceedings of the 38th International Conference on Software Engineering*, ser. ICSE '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 273–284. [Online]. Available: https://doi.org/10.1145/2884781.2884806

[10] F. Santos, B. Trinkenreich, J. a. F. Pimentel, I. Wiese, I. Steinmacher, A. Sarma, and M. A. Gerosa, "How to choose a task? mismatches in perspectives of newcomers and existing contributors," in *Proceedings of the 16th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement*, ser. ESEM '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 114–124. [Online]. Available: https://doi.org/10.1145/3544902.3546236

[11] W. Xiao, J. Li, H. He, R. Qiu, and M. Zhou, "Personalized first issue recommender for newcomers in open source projects," in *Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering*, ser. ASE '23. IEEE Press, 2024, p. 800–812. [Online]. Available: https://doi.org/10.1109/ASE56229.2023.00158

[12] I. Steinmacher, S. Balali, B. Trinkenreich, M. Guizani, D. Izquierdo-Cortazar, G. G. Cuevas Zambrano, M. A. Gerosa, and A. Sarma, "Being a mentor in open source projects," *Journal of Internet Services and Applications*, vol. 12, no. 1, p. 7, Sep 2021. [Online]. Available: https://doi.org/10.1186/s13174-021-00140-z

[13] S. Balali, U. Annamalai, H. S. Padala, B. Trinkenreich, M. A. Gerosa, I. Steinmacher, and A. Sarma, "Recommending tasks to newcomers in oss projects: How do mentors handle it?" in *Proceedings of the 16th International Symposium on Open Collaboration*, ser. OpenSym '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3412569.3412571

[14] X. Tan, Y. Chen, H. Wu, M. Zhou, and L. Zhang, "Is it enough to recommend tasks to newcomers? understanding mentoring on good first issues," in *Proceedings of the 45th International Conference on Software Engineering*, ser. ICSE '23. IEEE Press, 2023, p. 653–664. [Online]. Available: https://doi.org/10.1109/ICSE48619.2023.00064

[15] M. Guizani, T. Zimmermann, A. Sarma, and D. Ford, "Attracting and retaining oss contributors with a maintainer dashboard," in *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society*, ser. ICSE-SEIS '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 36–40. [Online]. Available: https://doi.org/10.1145/3510458.3513020

[16] S. Gazanchyan, "Awesome first pr opportunities," 2020, [Online]. Available: https://github.com/MunGell/awesome-for-beginners.

figures/method2.pdf

Figure 1. Overview of the study method.

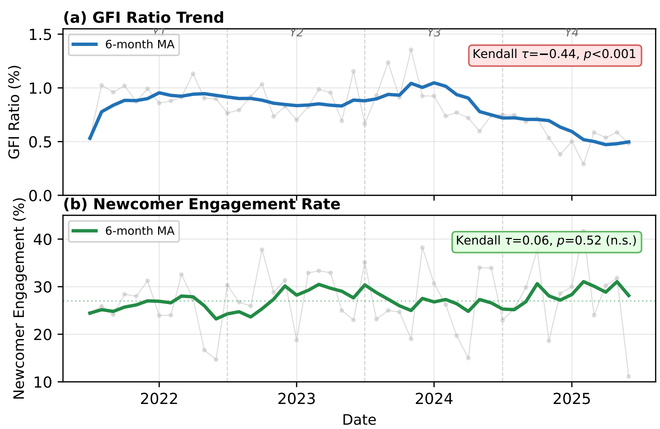

Figure 2. RQ1 time-series trends. (a) Monthly GFI ratio shows a significant decreasing trend ($\tau = -0.44$, $p < 0.001$). (b) Newcomer engagement rate remains stable at approximately 27% ($\tau = 0.06$, n.s.). Gray dots: monthly values; colored lines: 6-month moving averages; dashed vertical lines: analysis year boundaries.