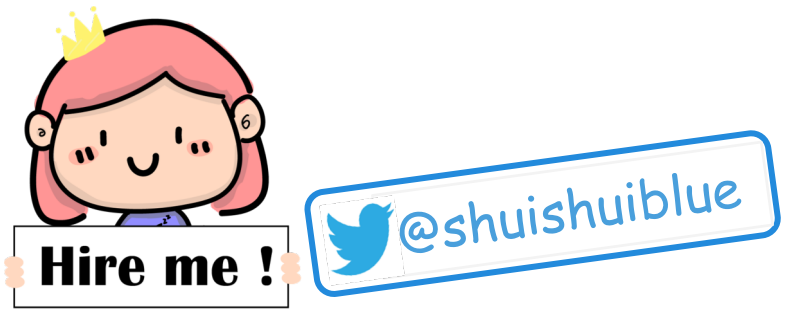
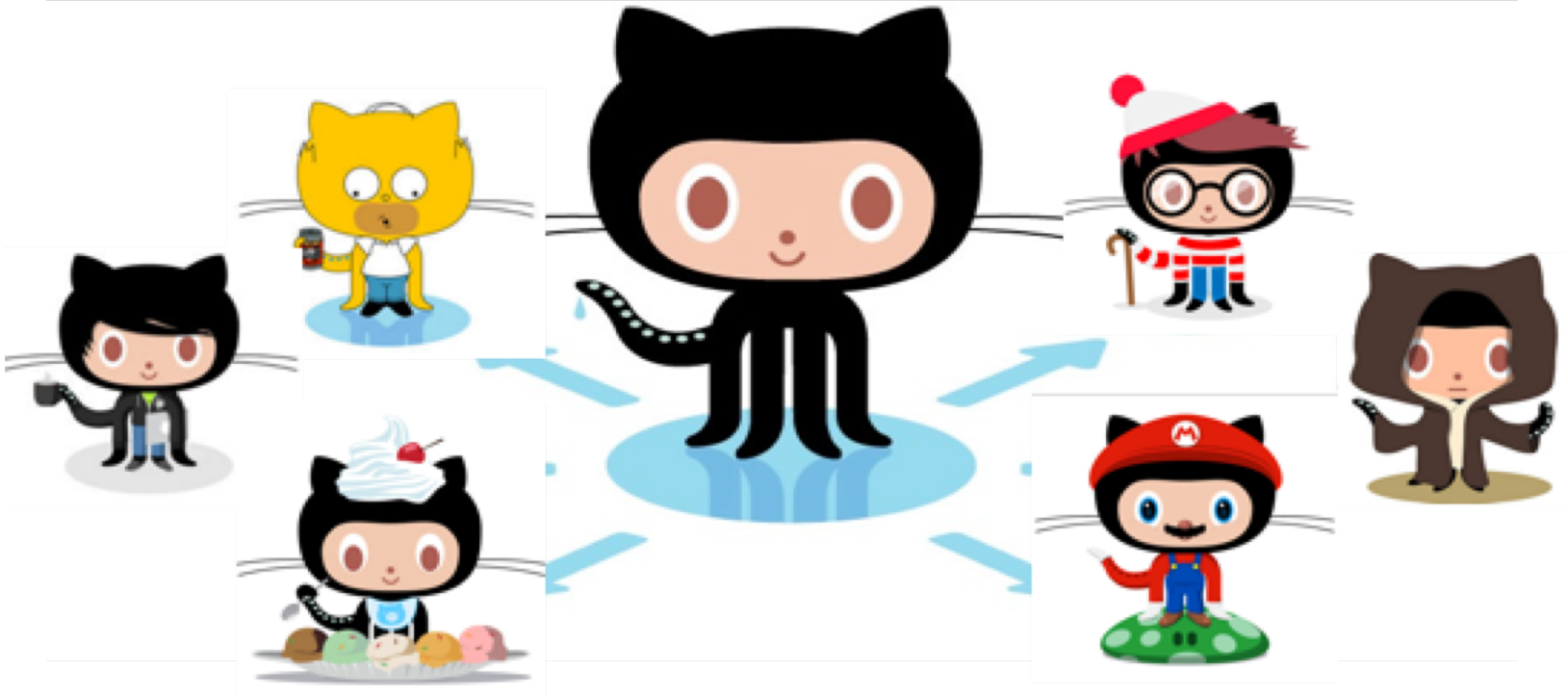
What the Fork: A Study of Inefficient and Efficient Forking Practices in Social Coding

Shurui Zhou, Bogdan Vasilescu, Christian Kästner



Carnegie Mellon University

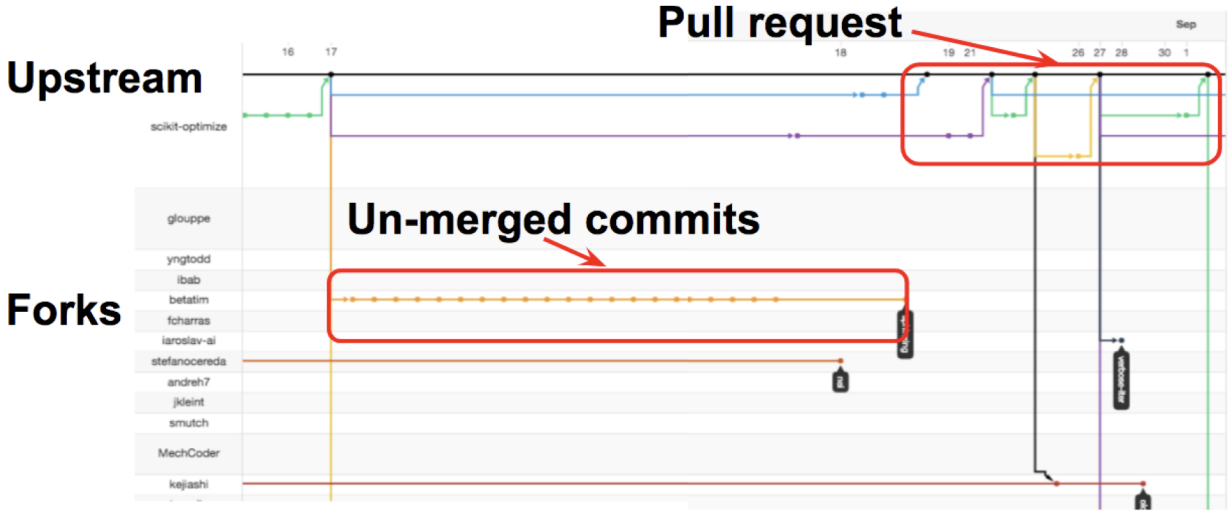
Fork-based Development



Fork-based Development is Popular

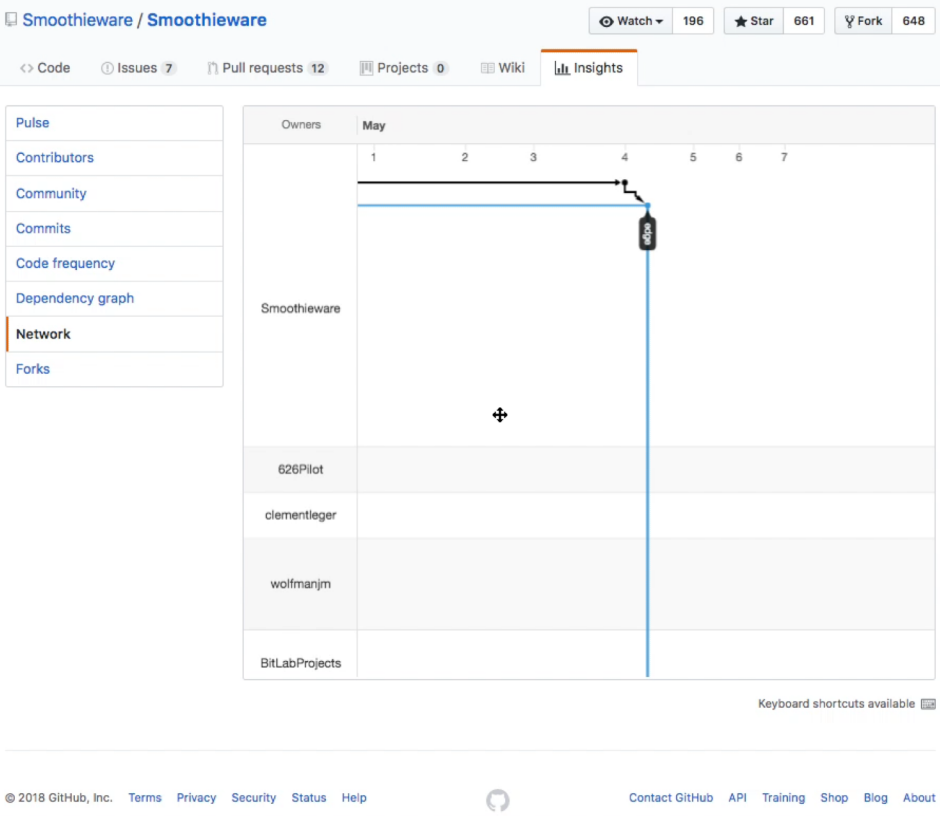
#Forks	#GitHub Projects
>50	61704
>500	4787
>1,000	2236
>5,000	198
>10,000	72
>100,000	2

[GHTorrent 2019-06]

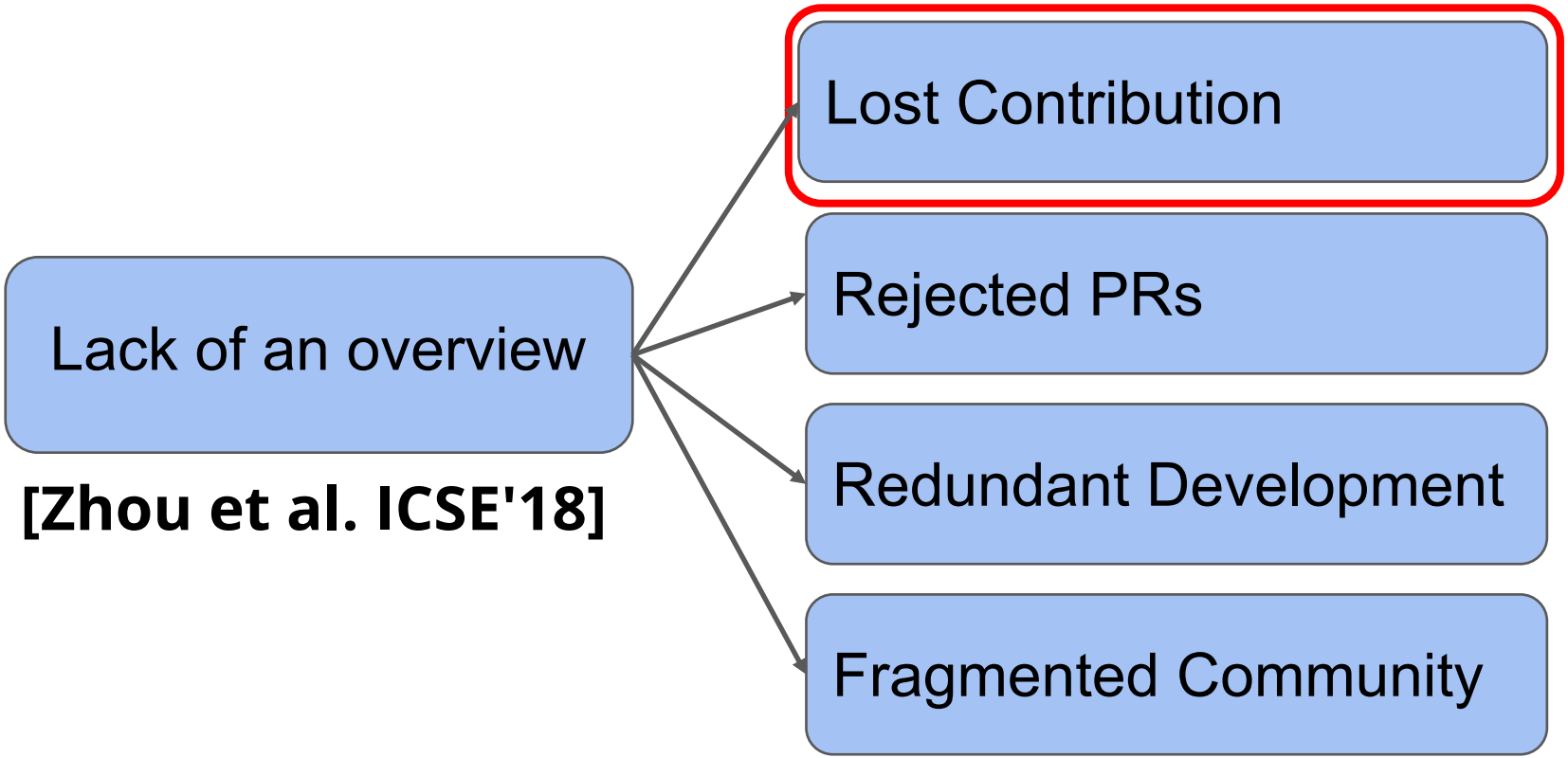


GitHub Network View

Network View - Lack of an overview



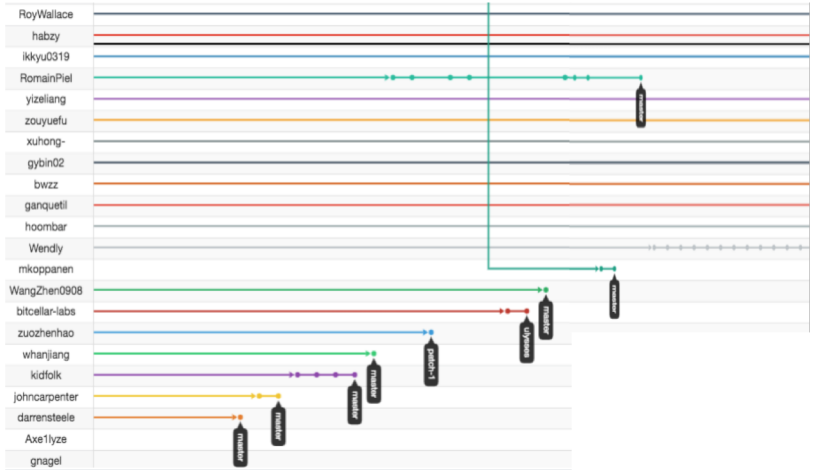
Problems ⟶ Inefficiency

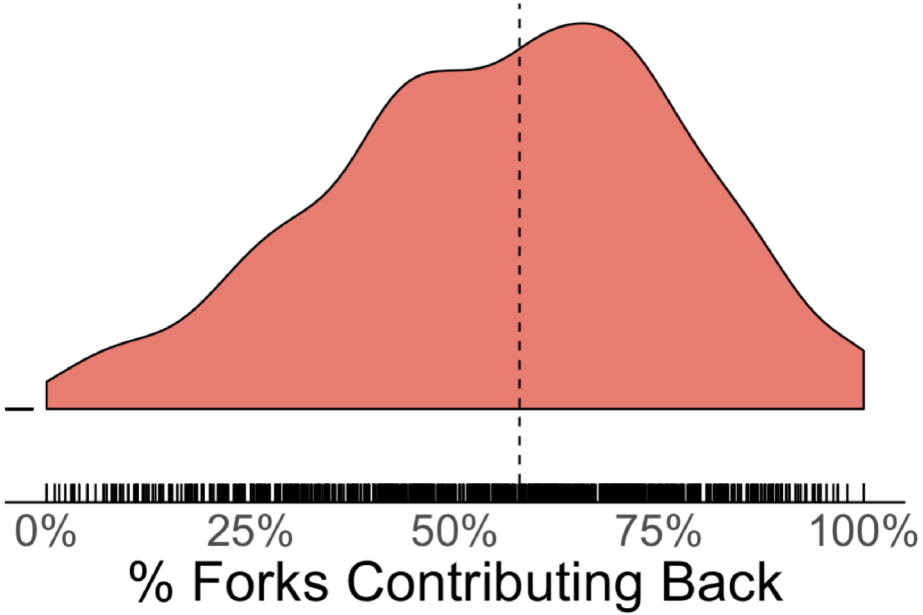


[Zhou et al. ICSE'18]

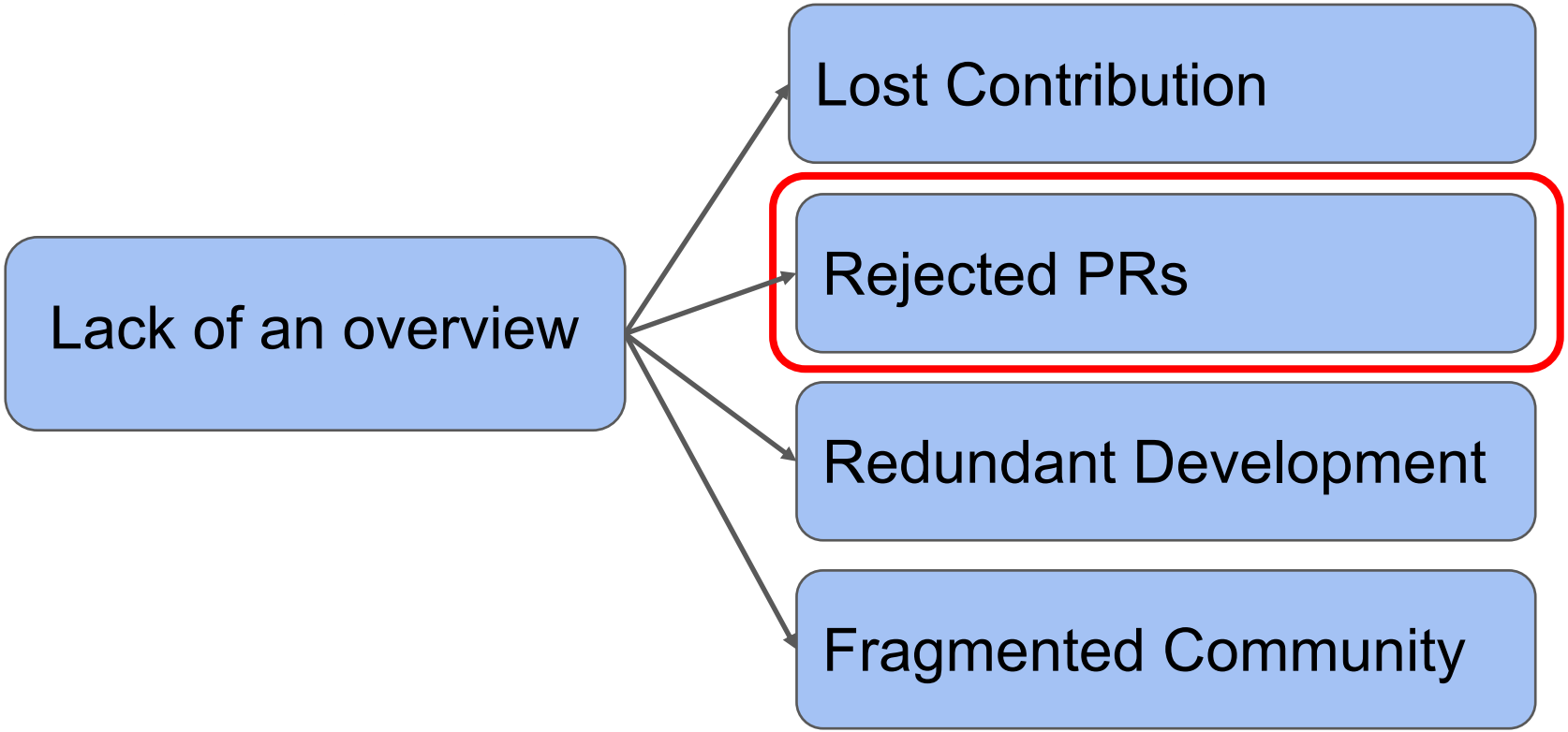
Lost Contribution

Only 14% of all forks of nine popular JavaScript projects on GitHub contained changes that were integrated back [Fung et al. 2012]





Problems ⟶ Inefficiency



Rejected Pull Requests

- Demotivating [Steinmacher et al. ICSE'18]
- Misalignment with maintainers' vision of the project

People Follow Different Processes

vs

People Follow Different Processes

"To a large extent the features are driven by bitcoin improvement proposals, so if I would be looking for a feature, I would go for these proposals"

--Bitcoin developer

People Follow Different Processes

Fix issue #13048 - Documentation regarding p-value bootstrapping #14759

Closed **achievermina** wants to merge 7 commits into `scikit-learn:master` from `achievermina:p_valueBootstrapping`

Conversation 9 | Commits 7 | Checks 11 | Files changed 2

achievermina commented 3 days ago • edited

Issue #13048

People Follow Different Processes

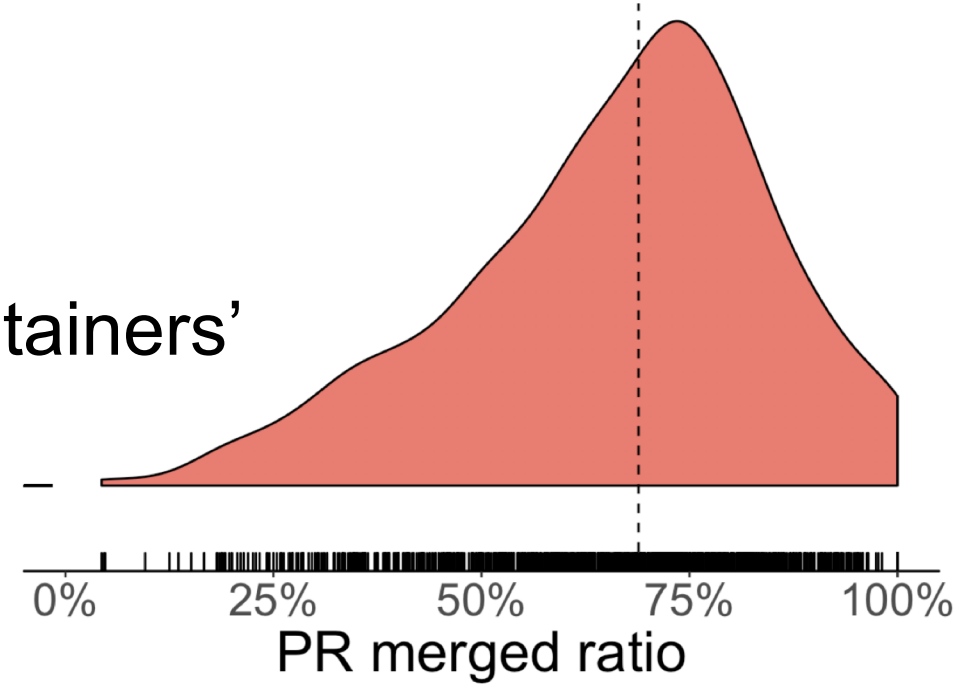
vs

- Project proposal
- Resolve issues on the issue tracker

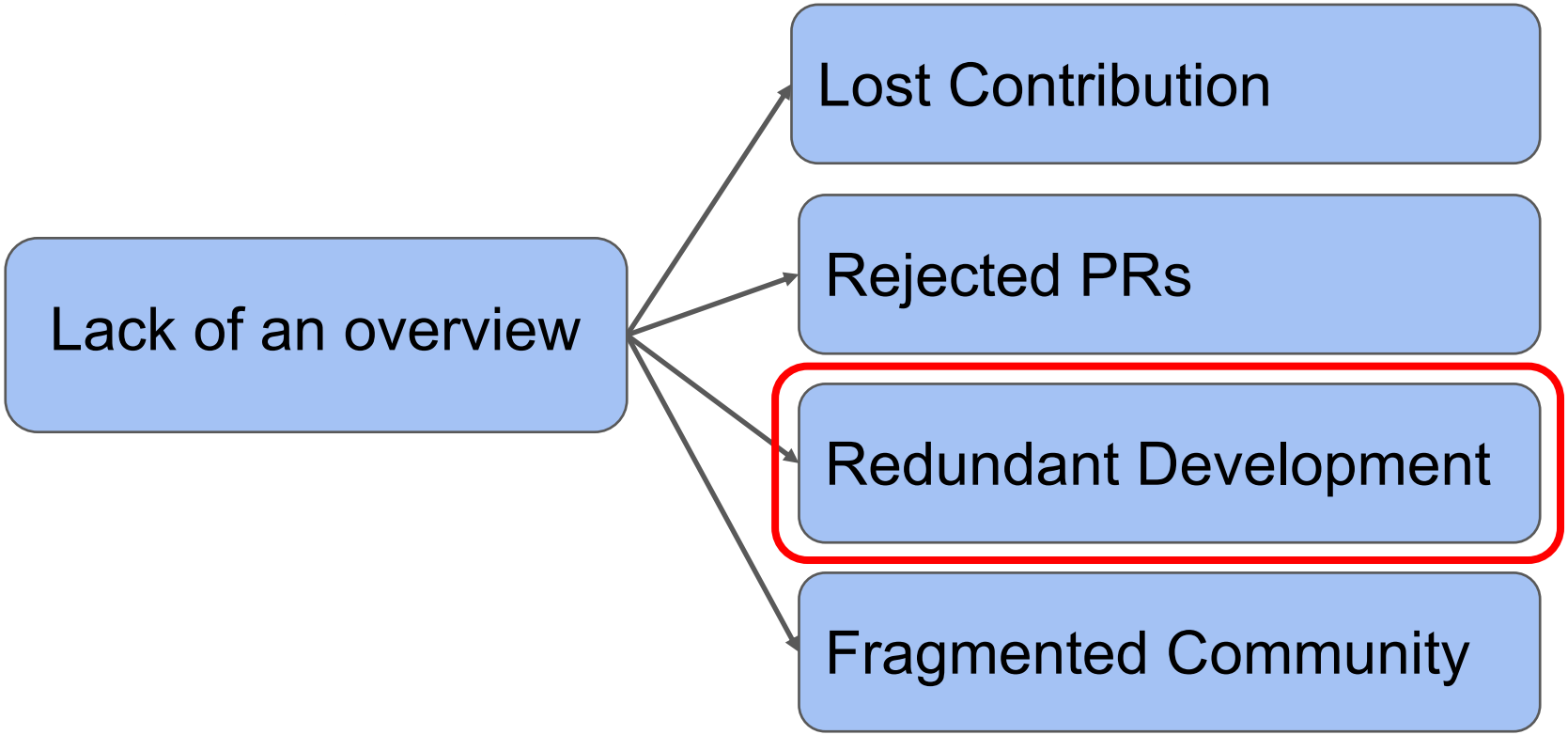
- Open for any contribution

Rejected Pull Requests

- Demotivating
- Misalignment with maintainers' vision of the project



Problems ⟶ Inefficiency



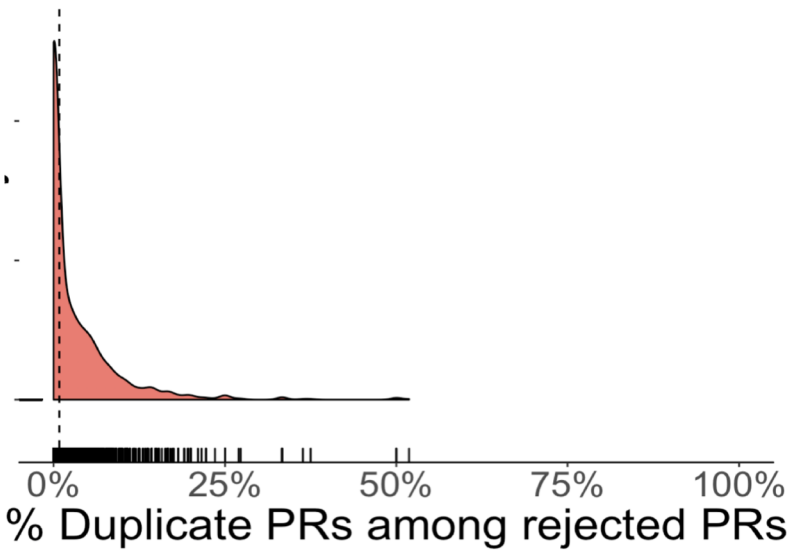
Redundant Development

23% un-merged PRs were rejected due to redundant dev. [Gousios et al. ICSE'14]

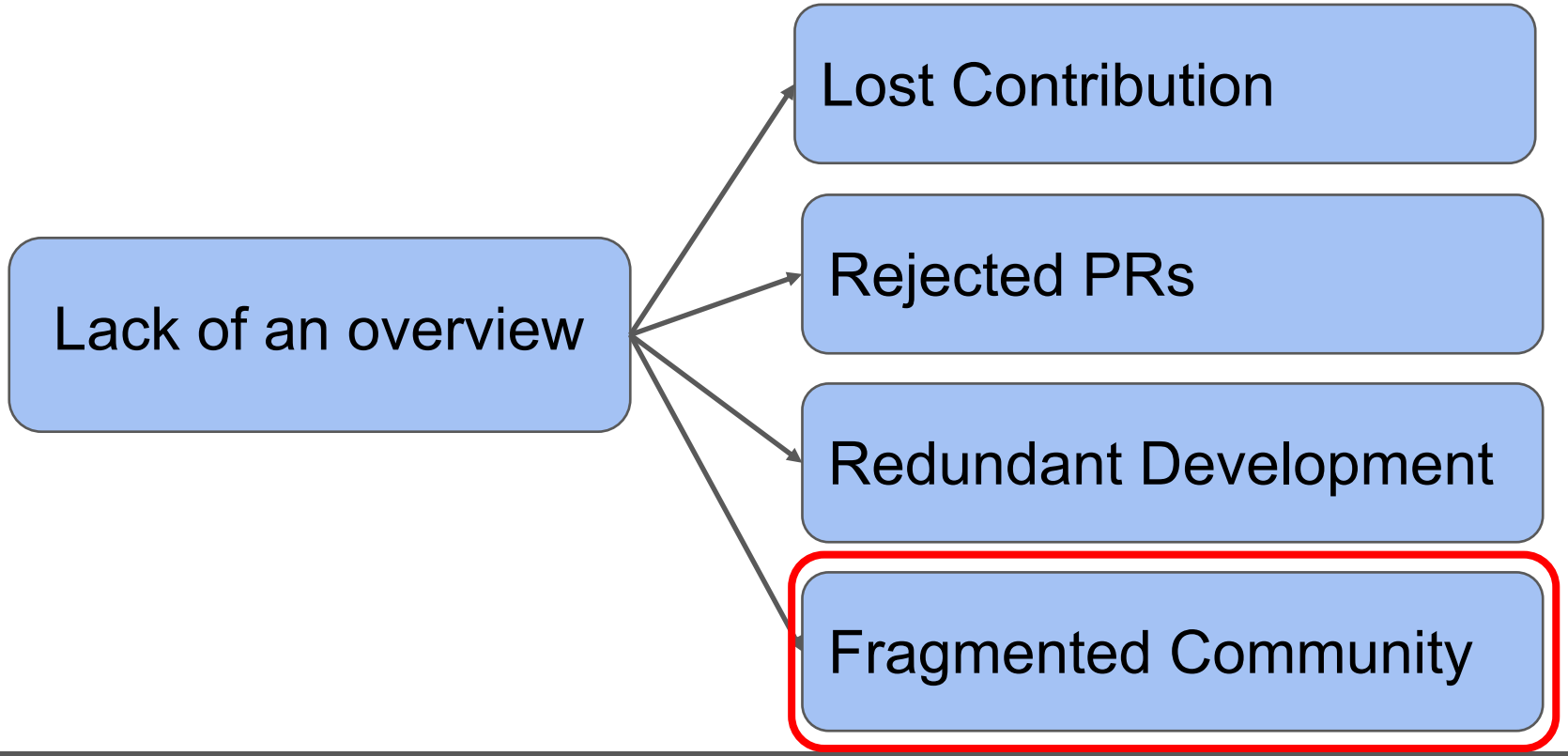
Cost of Reviewing [Li et al. MSR'18]

De-motivate developers [Steinmacher et al. ICSE'18]

Detecting duplicate dev. [Zhou et al. SANER'19]



Problems ⟶ Inefficiency

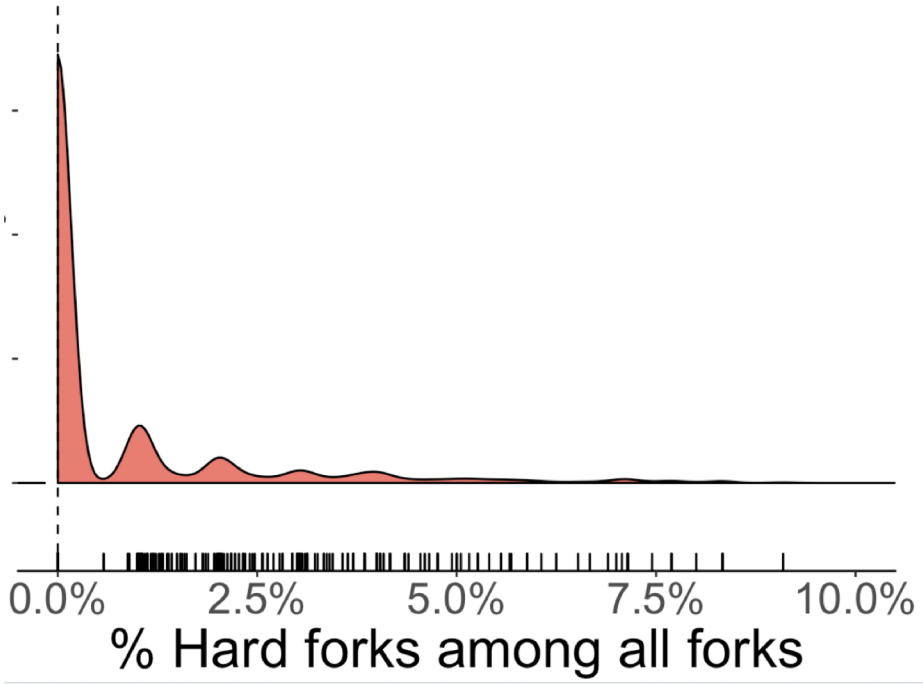


Communities Fragmentation (Hard Fork)

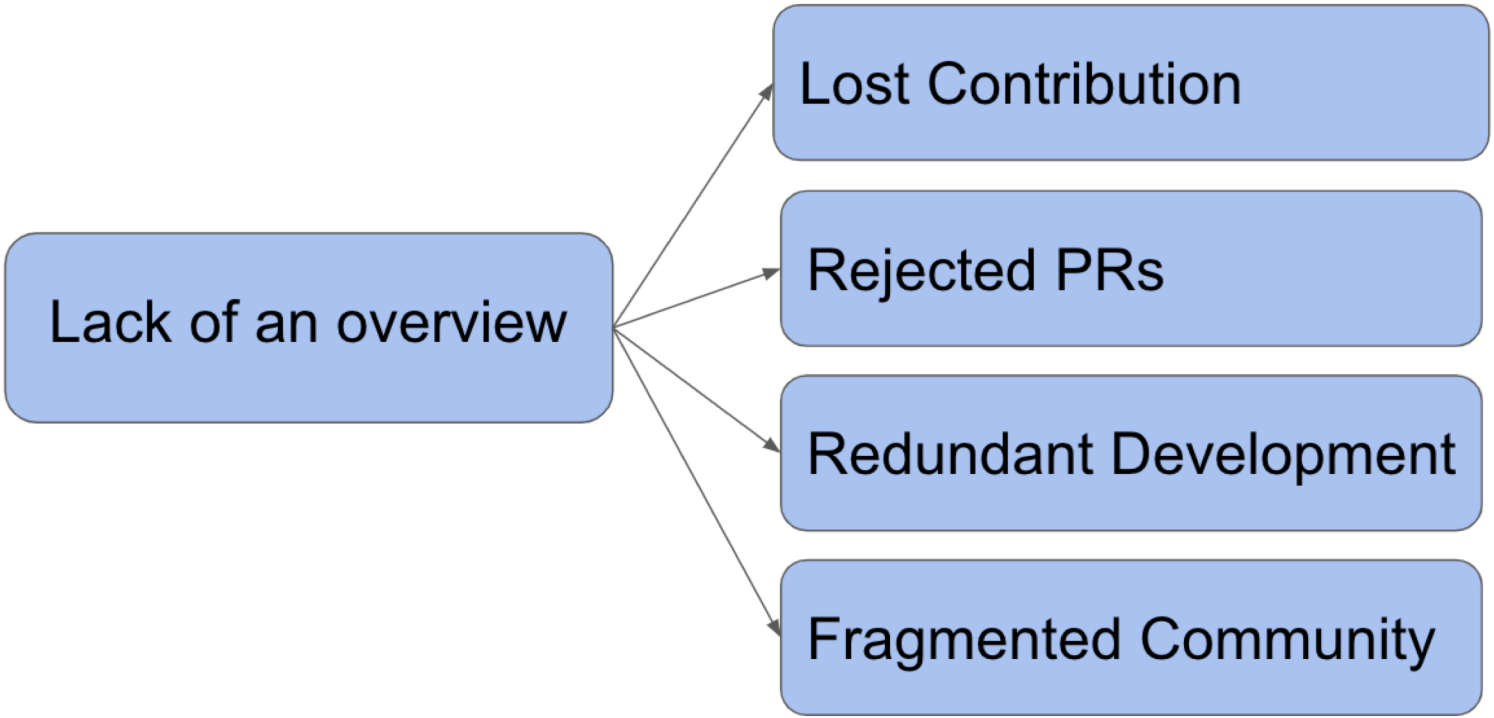
Behind the Scenes Bytes

3D Printer Firmware – Which to Choose and How to Change It?

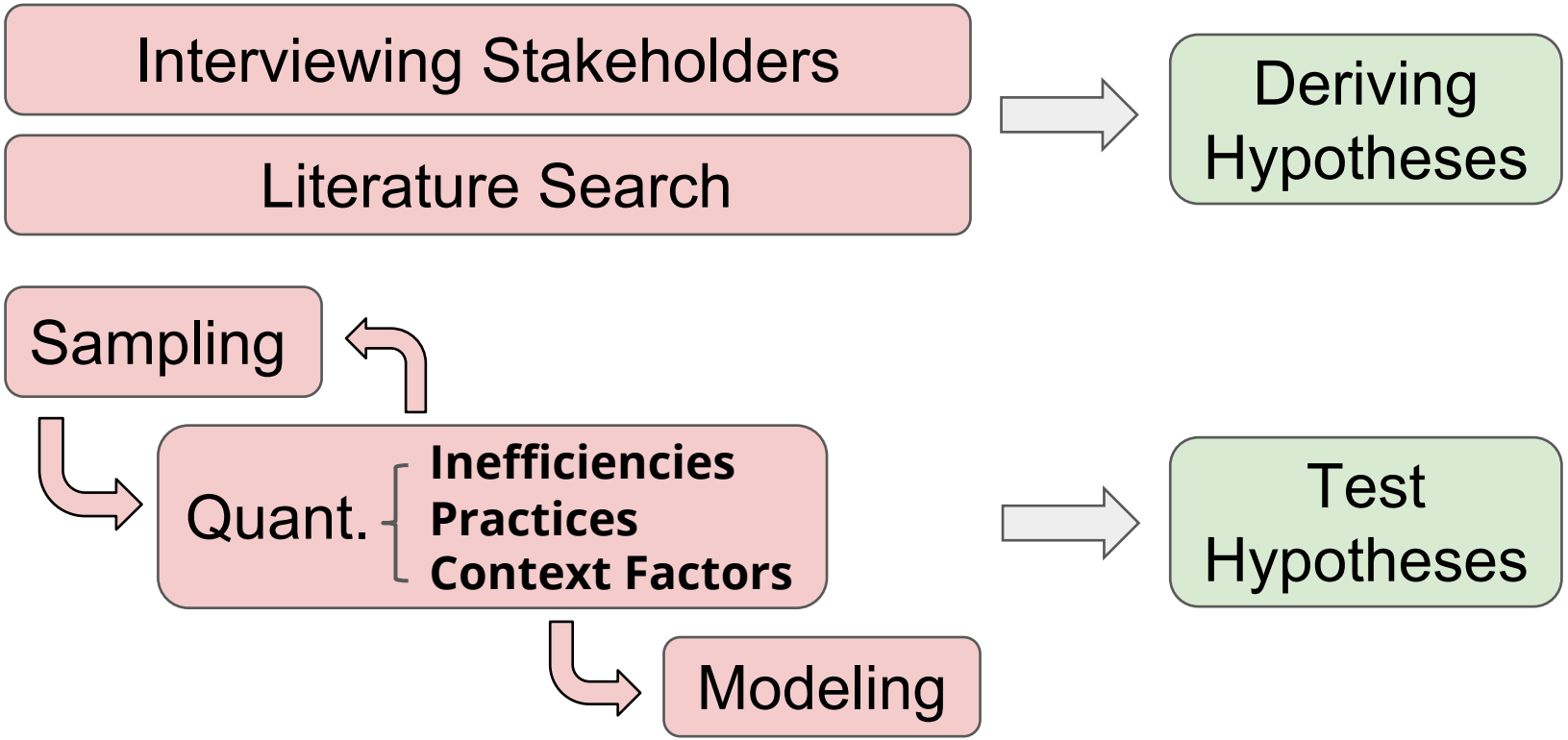
by *Michael Jones*
Apr 4, 2018



RQ: What characteristics and practices of a project associate with efficient forking



Research Method



Coordination Mechanism Affects Forking Practices

vs

- Project proposal
- Resolve issues on the issue tracker

- Open for any contribution

Coordination Mechanism Affects Forking Practices

Centralization makes it easier to coordinate the divisions' product types but more difficult to take advantage of the divisions' private information. [Brandts et al. 2018]

Deriving Hypotheses

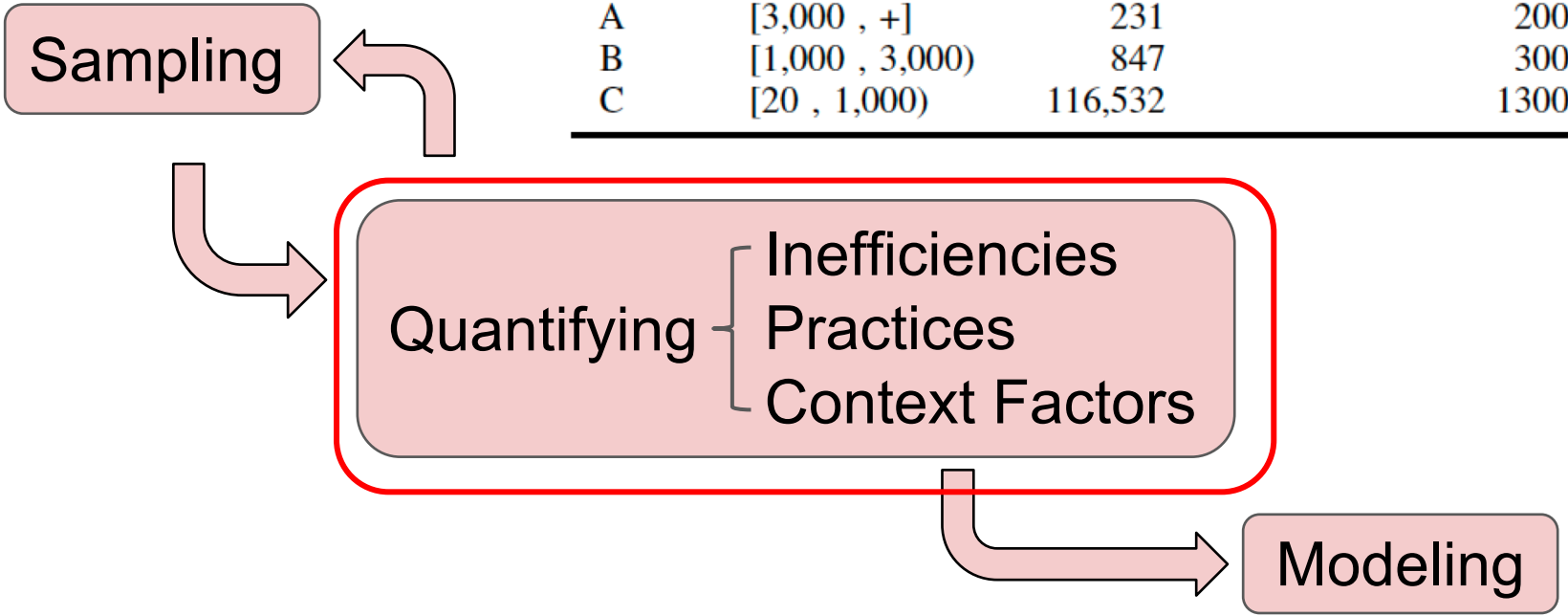
Centralized mgmt ➔ Larger portion of merged PRs

Centralized mgmt ➔ Larger portion of contributing forks

(6 more in the paper)

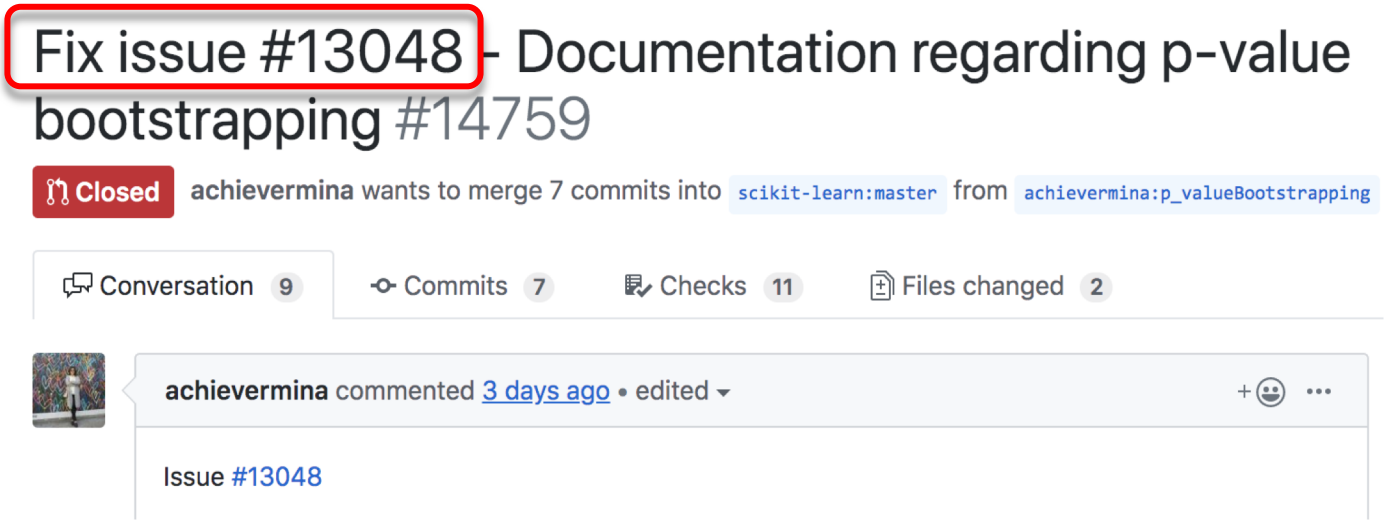
Test Hypotheses

Group	#fork	#projects	#projects in sample set
A	[3,000 , +]	231	200
B	[1,000 , 3,000)	847	300
C	[20 , 1,000)	116,532	1300

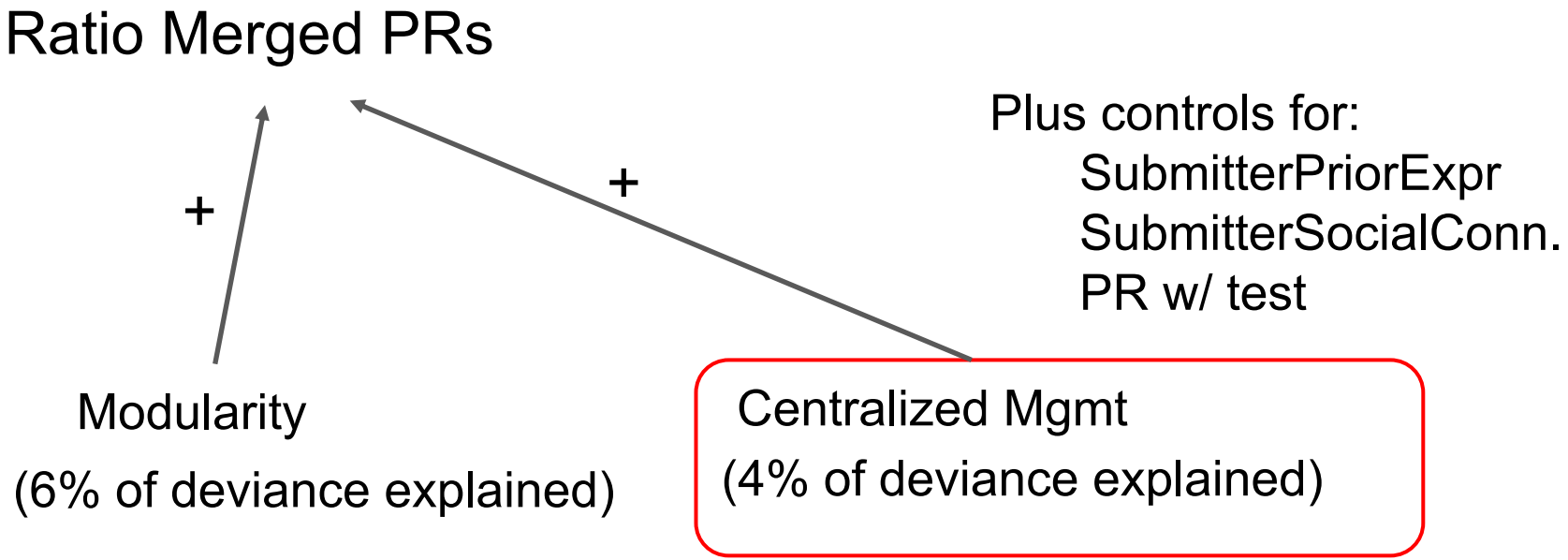


Operationalization - Centralized Management

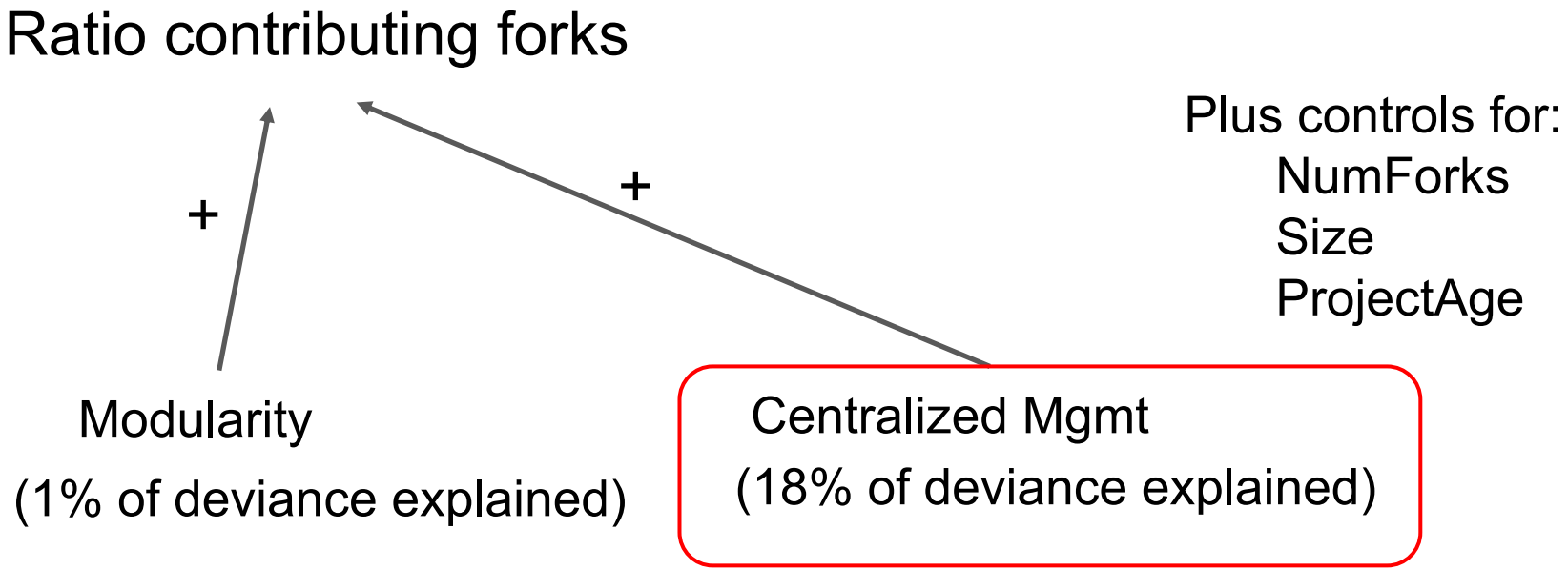
Measure: $\frac{\text{Number of PRs referring to an Existing Issue}}{\text{All the PRs}}$



Centralized Mgmt → More Merged PRs (R^2 = 27%)



Centralized Mgmt → More Contributing Forks (R^2 = 17%)



Evidence-based Intervention

For practitioners:

- Coordinating planned changes through an issue tracker

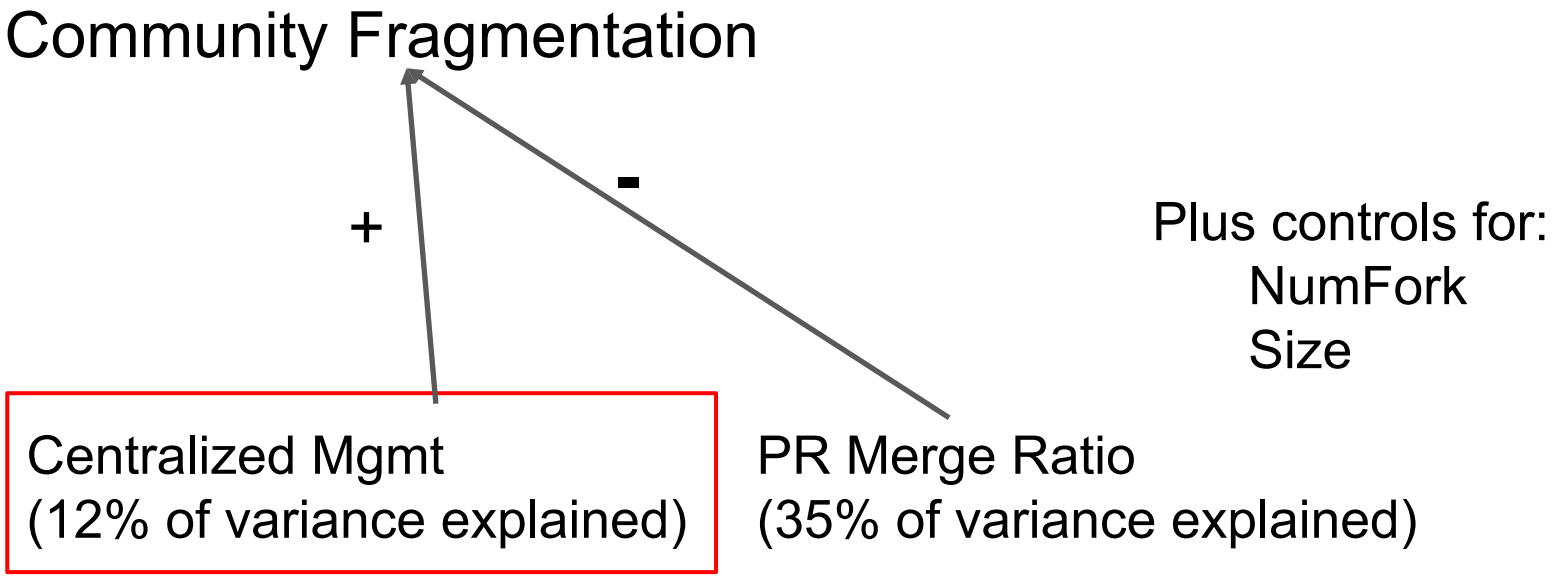
Trade-offs?

- Project proposal
- Resolve issues on the issue tracker

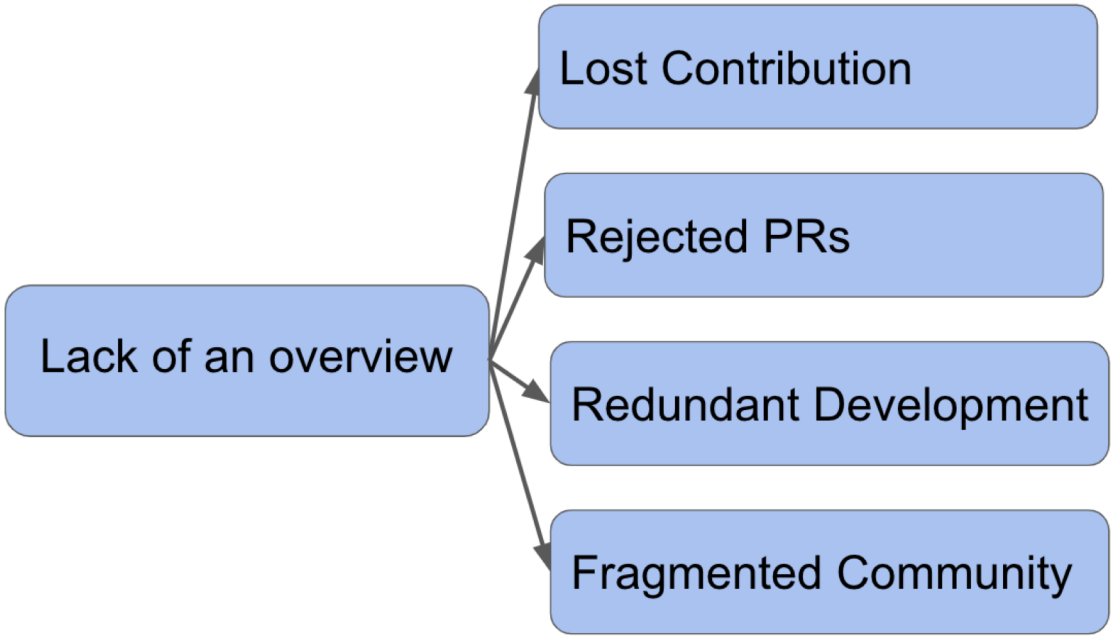
VS

- Open for any contribution

Trade-off: Centralized Mgmt



RQ: What characteristics and practices of a project associate with efficient forking practices?



- Coordination
- Modularity

Opportunities to Design Further Interventions

- Tooling to navigate and understand changes in forks
- Making practices transparent

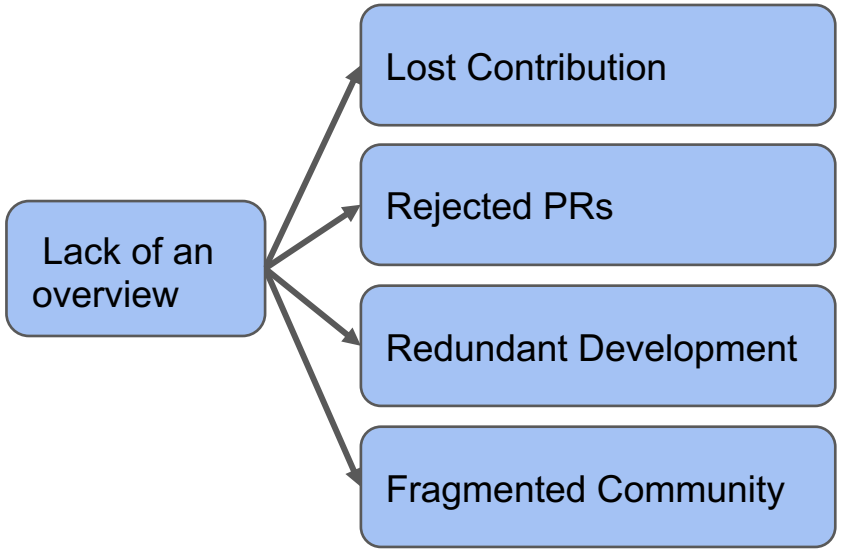
- Cost of community fragmentation

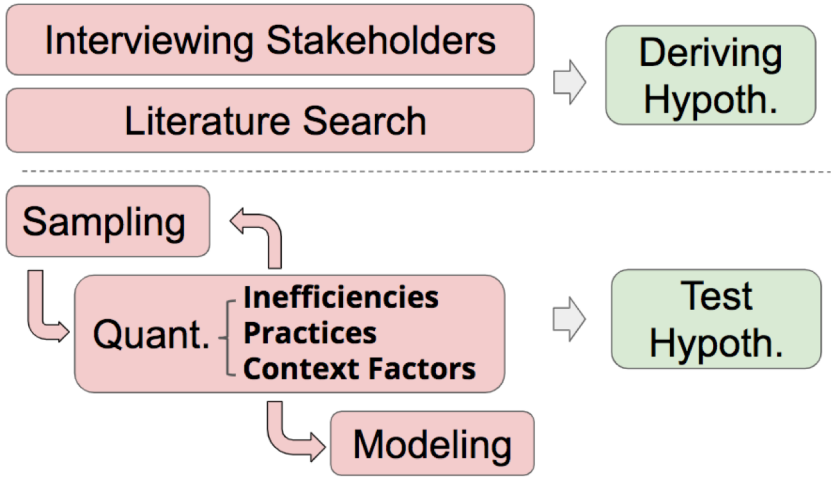
No Spoons

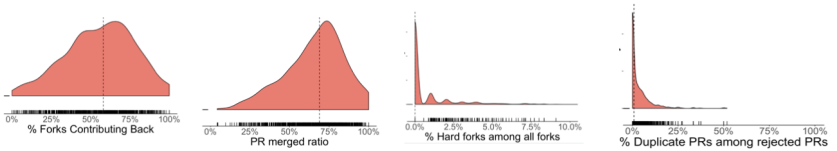
NO KNIFE

A Study of Inefficient and Efficient Forking Practices in Social Coding

@shuishuiblue







- **Evidence-based Suggestions**
- **Further research/tooling directions**