



# Healthy Open-Source Communities: What Does The Empirical Evidence Say

Bogdan Vasilescu  
Wednesday, April 2nd 2025

# The Internet is full of tips

<https://github.blog/open-source/>

**Top tips for creating a healthy and sustainable open source community**

Read about the six key themes, and tips for each.

## What makes a sustainable open source community?

The six key areas that are characteristic of communities are that they:

- Have a **README** file
- Have **good-first-issues**
- Contain **contribution guidelines**
- Contain a **code of conduct**
- Encourage respectful language in GitHub
- Commit to **mentorship**

**Red Hat**

## Checklist for measuring the health of an open source project

A [healthy open source project](#) demonstrates open practices, uses open infrastructure, has an open culture with the goal of becoming more sustainable. Use this checklist to assess the health of your open source project. The more checks you see when you are finished, the healthier your project will be.

### Infrastructure

**Basic infrastructure\***

- Code repositories and [issue trackers](#) are present and easily accessible
- Development and testing tools are present and easily accessible
- Project website is present and easily accessible
- Mailing list servers are present and easily accessible
- Documentation platform is present and easily accessible
- Video conferencing platform is present and easily accessible
- Community forum is present and easily accessible
- Synchronous chat tools are present and easily accessible
- Community event calendar is present and easily accessible

\* Infrastructure is "easily accessible" if users and contributors do not require special permission or specific institutional affiliation to engage with it.

### Project website

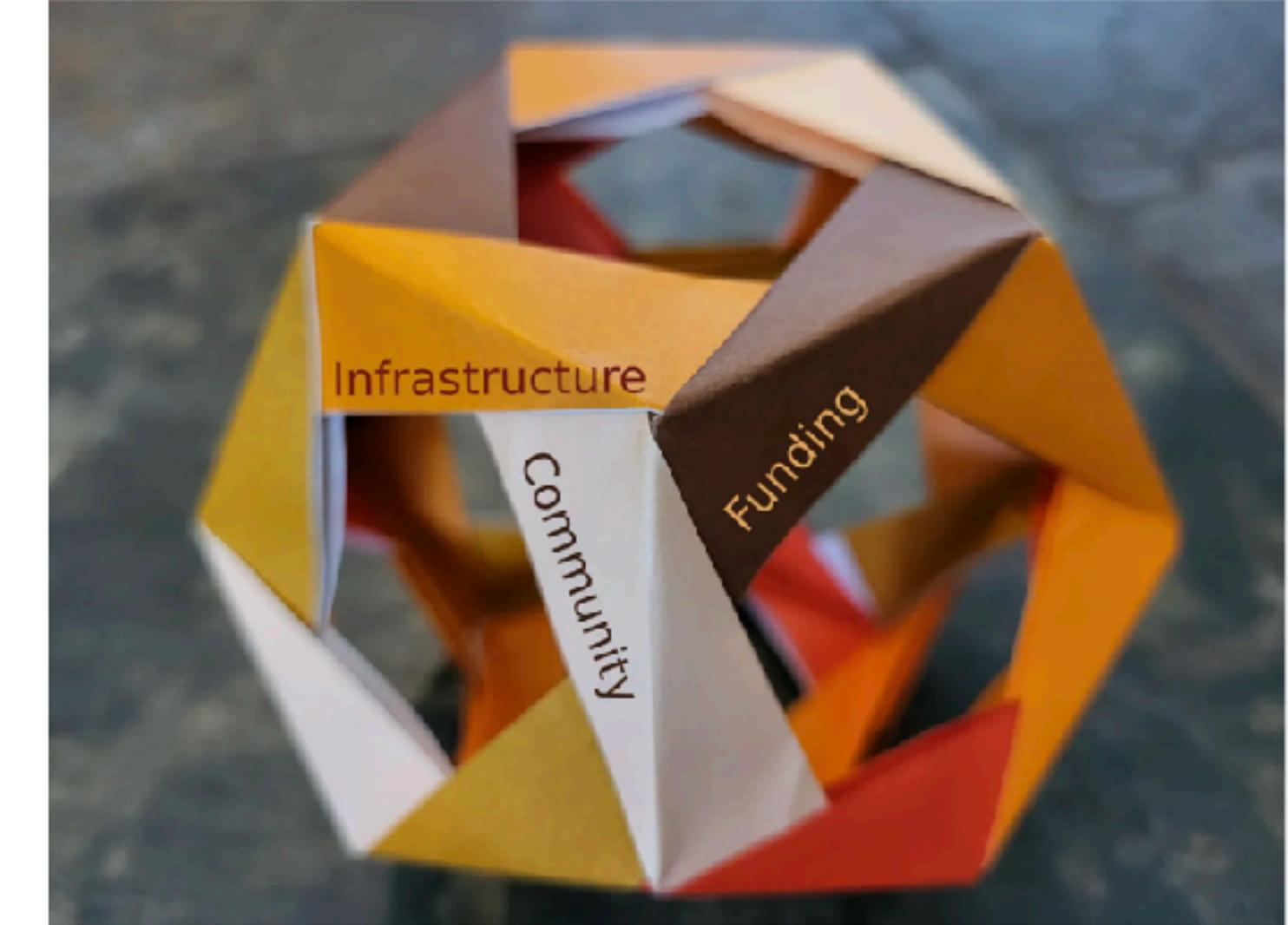
- Website has a unique domain name
- Website features other publicly visible sites or subdomains

<https://medium.com/spiffworkflow/how-we-maintain-a-healthy-open-source-project-2e6d7115f668>

**How we maintain a healthy open source project**

Dan Funk · [Follow](#)  
Published in SpiffWorkflow · 11 min read · Apr 15, 2024

53 · Q



Modular origami has always felt a little like software development — fitting reusable components together.

# Science is needed for evidence-based recommendations

## Anecdotal evidence reliable? One man says “yes”.

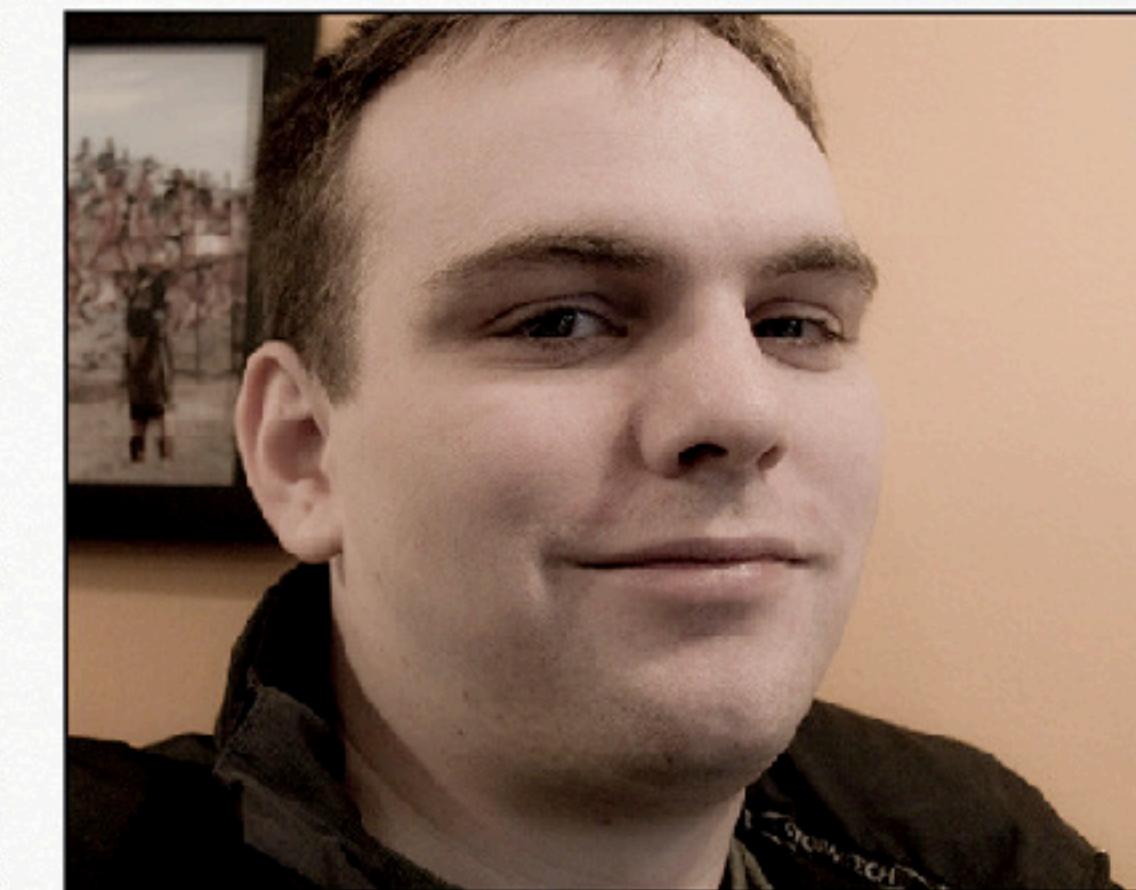
A STUDY CONDUCTED YESTERDAY by a man on himself concluded that self-reported anecdotal evidence is, in fact, both reliable and relevant.

The landmark study, conducted by Mark Mattingly of Virginia Beach in his apartment, concluded with 100% accuracy that data collected from personal experience can disprove other data conducted by reputable scientific institutions, thereby proving once and for all that “statistics can’t be trusted”.

In a press release Mr. Mattingly took aim at his detractors saying that “...this study shows what I’ve been telling people on the internet for years: all your fancy evidence and statistics don’t mean nothing in the real world.”

A frequenter of internet forums, comment sections, and social media, Mr. Mattingly recounts that he was inspired to undertake the study when someone reportedly kept insisting that he provide evidence for his claims. “I think everyone’s entitled to an opinion, and that my opinion is worth just as much as anyone else’s” Mr. Mattingly said.

Academic types have criticised the study, and papers who are publishing it, saying that it lacks everything and makes no sense. When shown the study, Emeritus Professor James Albrecht of Carnegie Mellon University looked all confused and hopeless before making pining, guttural sounds.



Mr. Mattingly in his apartment looking all smug

Mr. Mattingly has responded saying that this is just the first of many studies he intends to conduct, and that a meta-analysis of people who have opinions and anecdotal experiences independent of controls, methodological rigor, blinding and peer review are soon to be published, adding further weight to his initial findings.

# STRUDEL sustainability research on ...

## Project practices

- [CHASE 2023](#) (social media)
- [ICSE 2020](#) (forking)
- [ESEC/FSE 2019](#) (forking)
- [ESEC/FSE 2018](#) (abandonment factors)

## Funding models

- [ICSE 2020](#) (donations)

## Sunsetting

- [ESEC/FSE 2023](#) (dealing with abandonment)
- [ICSE 2025](#) (quantifying abandonment)

## Attracting contributors

- [ICSE 2022](#) (Twitter)
- [MSR 2020](#) (Twitter)
- [CSCW 2019](#) (signals)
- [ESEC/FSE 2015](#) (social connections)

## Stress, burnout, disengagement

- [ICSE 2022](#) (toxicity theory)
- [ICSE SEIS 2022](#) (toxicity vs pushback)
- [ICSE NIER 2020](#) (toxic language)
- [ICSE 2019](#) (overwork)
- [OSS 2019](#) (dropout, survival analysis)

## Transparency and signaling

- [ESEC/FSE 2020](#) (diffusion of practices)
- [CSCW 2019](#) (signals)
- [ICSE 2018](#) (badges)

## Diversity and inclusion

- [CHI 2023](#) (ClimateCoach)
- [ICSE SEIS 2023](#) (census)
- [ICSE 2019](#) (social capital)
- [CHI 2015](#) (gender & tenure)
- [CHASE 2015](#) (survey)

## Novelty and innovation

- [ICSE 2024](#) (atypical combinations)

## Network effects

- [ICSE 2024](#) (innovation)
- [ESEC/FSE 2023](#) (labor pools)
- [ICSE 2022](#) (Twitter)
- [ESEC/FSE 2020](#) (diffusion of practices)
- [ICSE 2019](#) (social capital)
- [ESEC/FSE 2018](#) (abandonment factors)

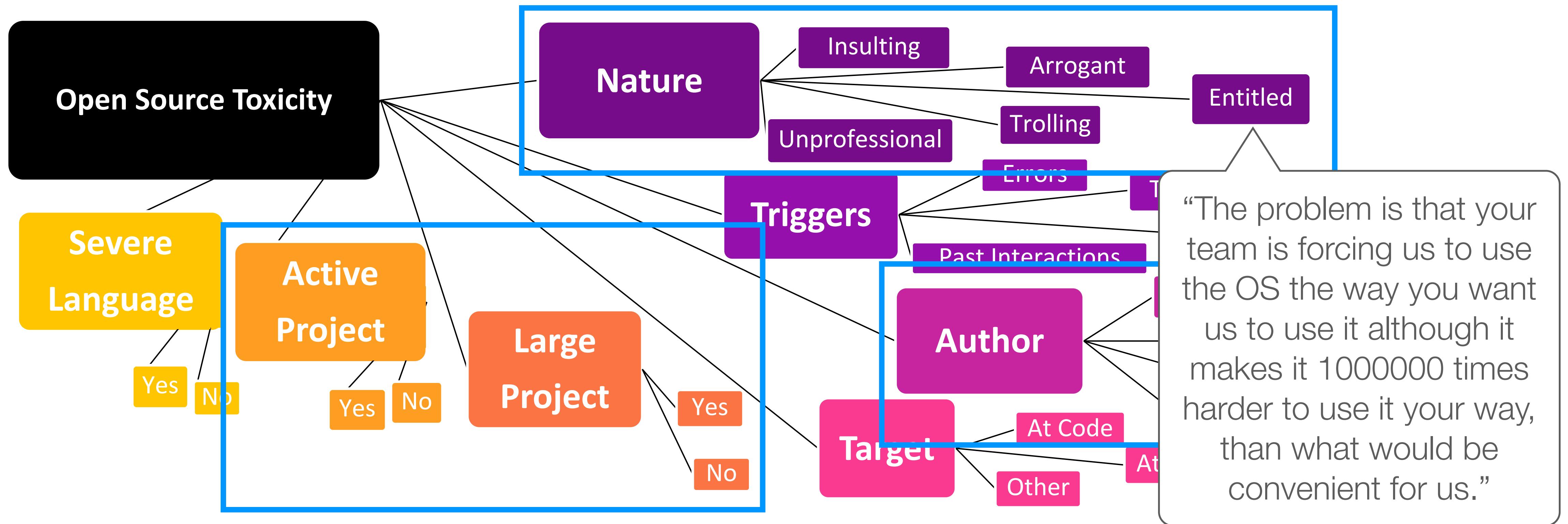
# Today: Open source ecosystem as ...

---

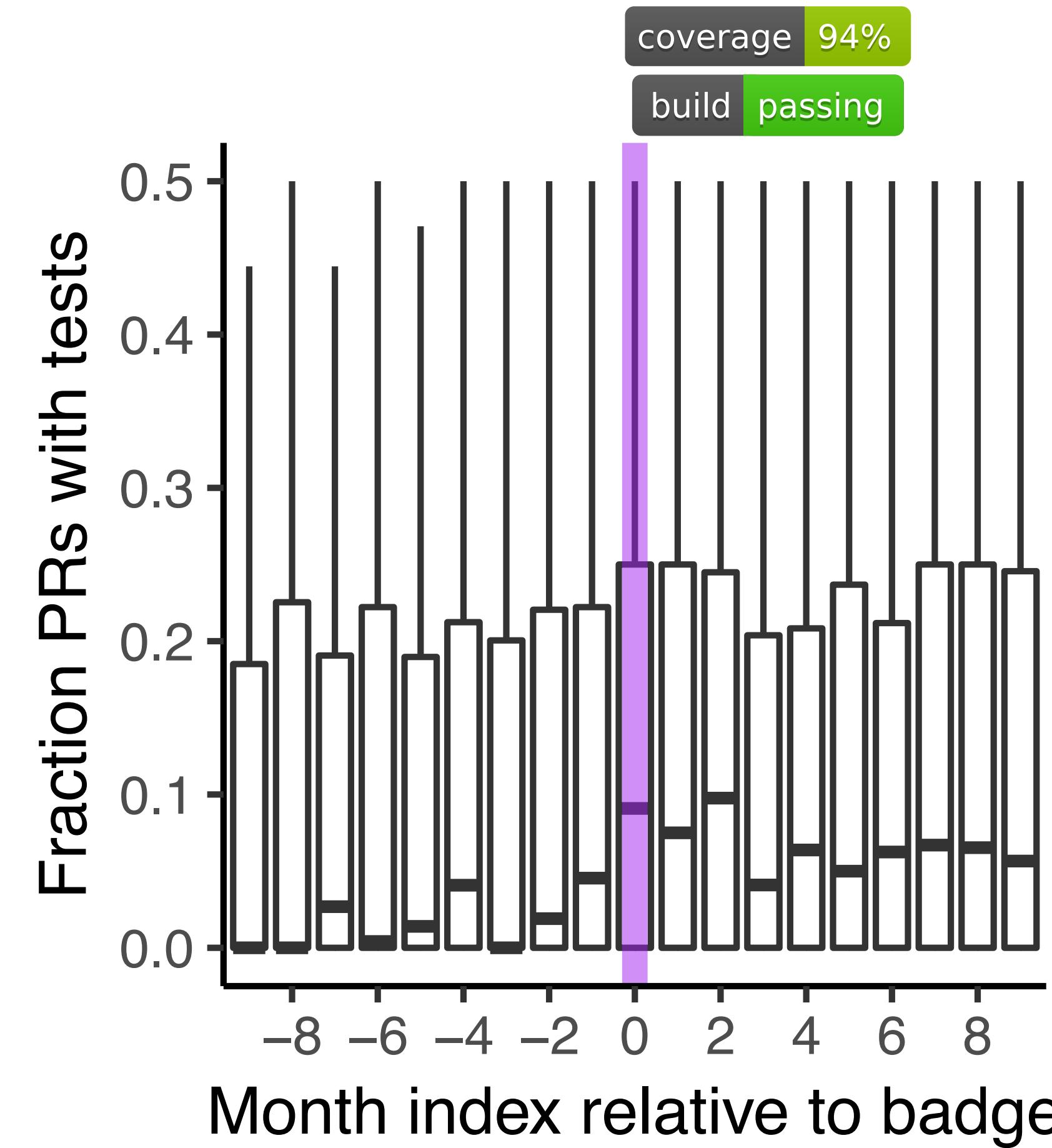
- Highly **transparent** environment
- **Attention** economy
- Socio-technical **network**

# Entitled comments are a frequent source of toxicity\*

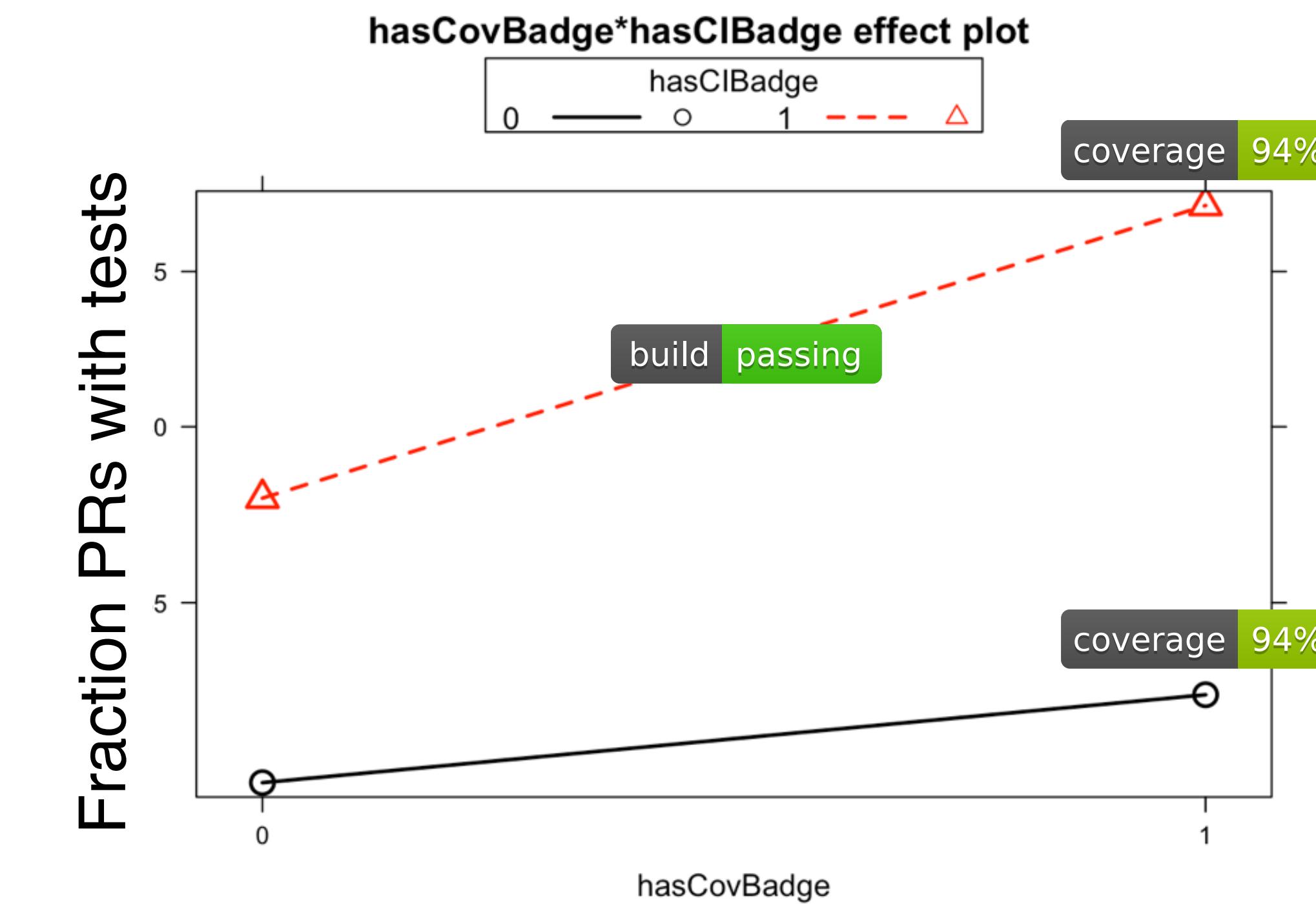
\*rude, disrespectful, or unreasonable comments that are likely to make someone leave a discussion (Google)



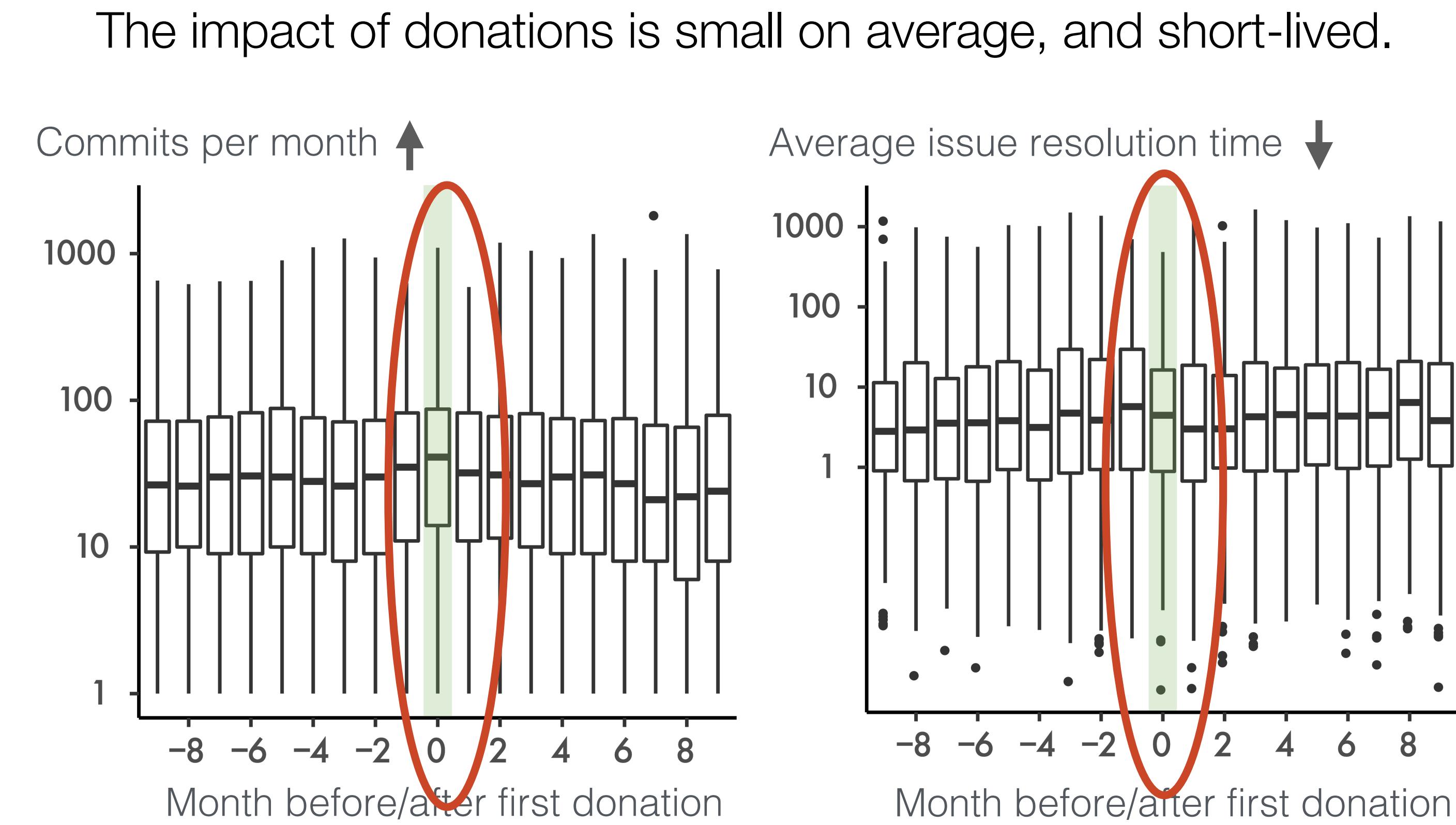
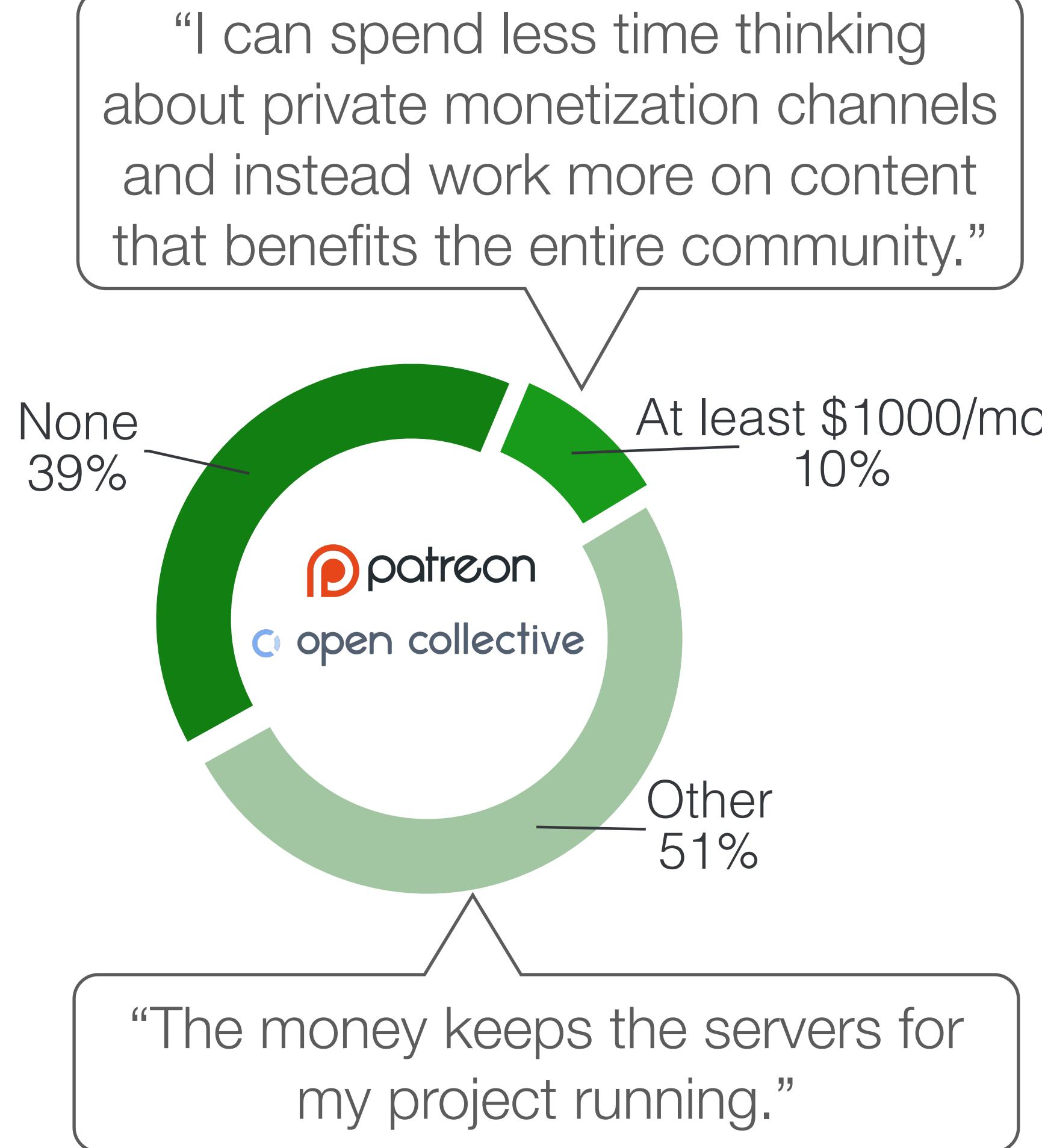
# The monthly fraction of PRs containing tests increases after adopting quality assurance badges



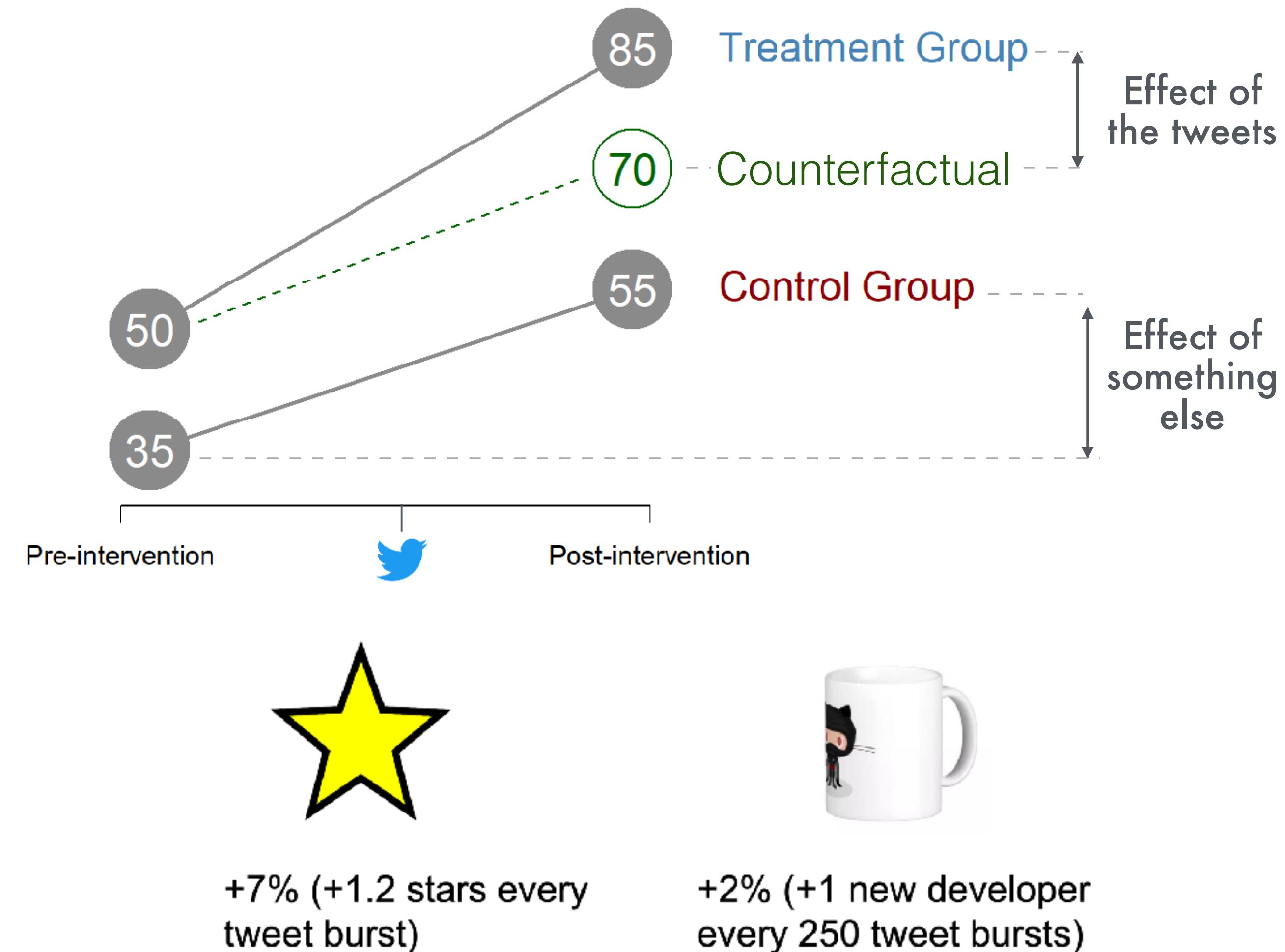
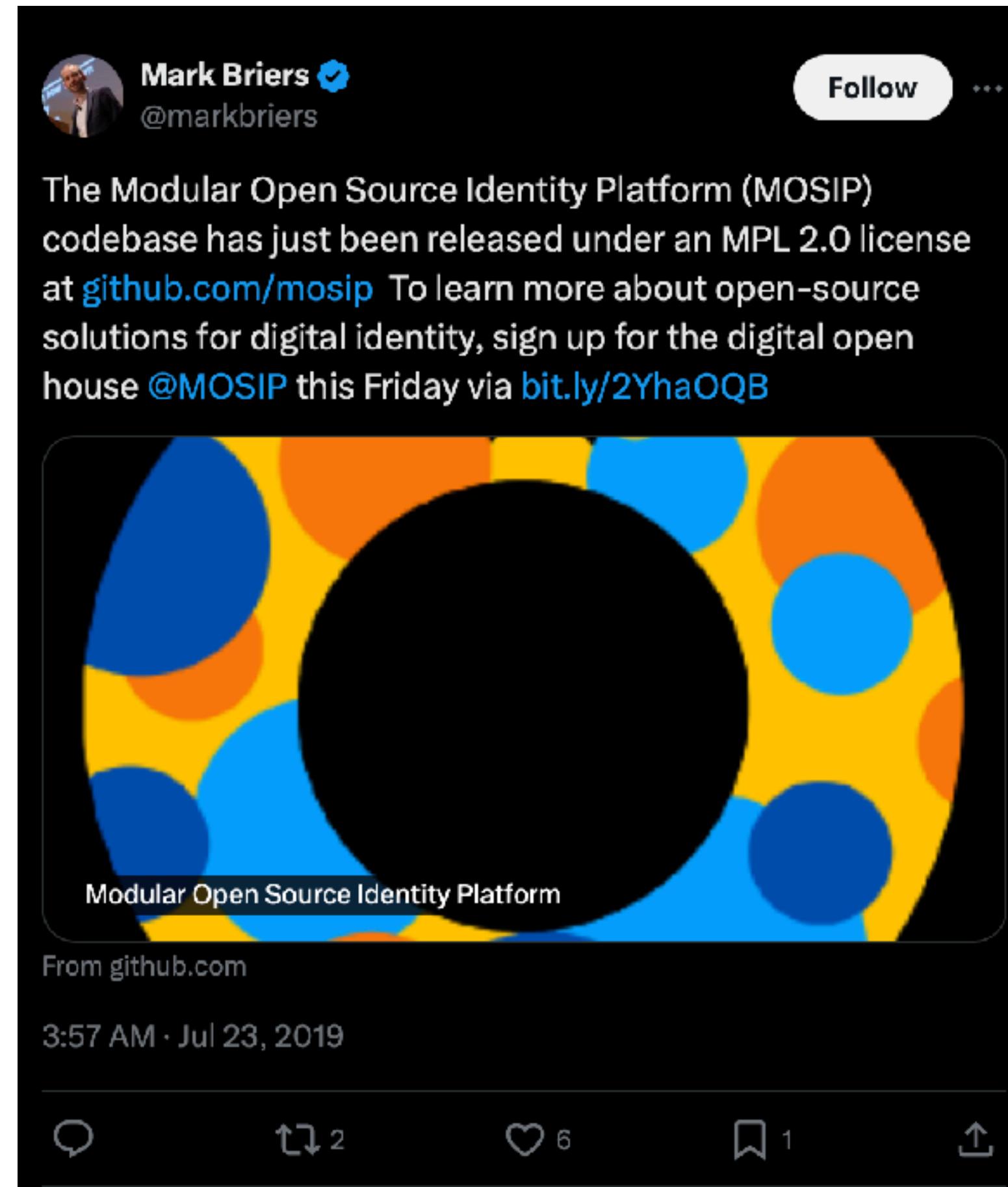
Coverage and CI badges interact, amplifying each other's effects.



# Crowdfunding helps with some things, but not enough



# Tweets attract new GitHub stars and contributors

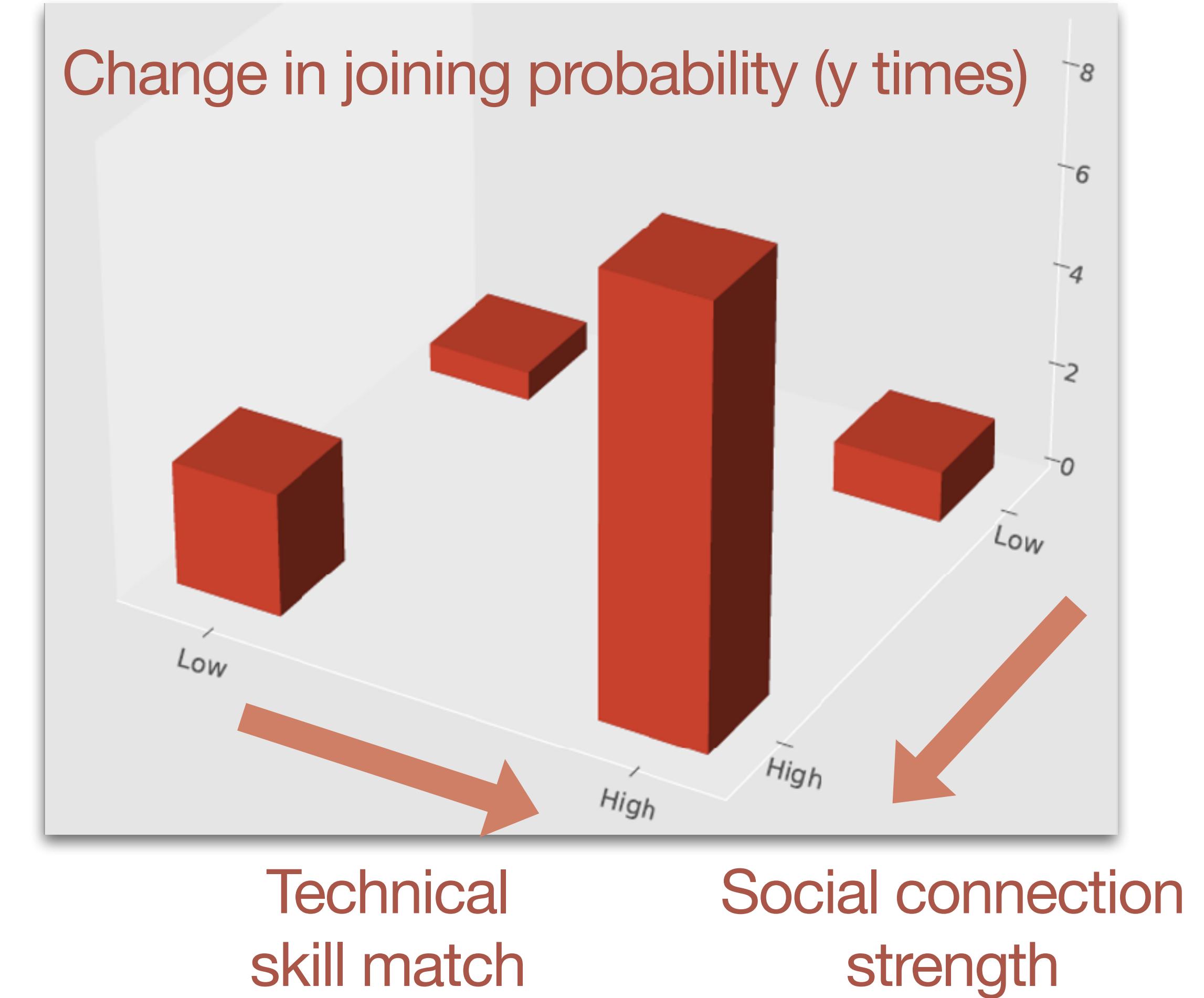


# Social connection strength, technical skill match, and amount of competition all explain variance in new contributors joining.

27% more variance explained by model with network effects vs only project-level characteristics

Individual-level effects (bottom 50% vs top 50%)

- Social connection strength ..... 6.95 x
- Technical skill match ..... 3.20 x
- Competition ..... -2.40 x



# Network ties predict the spread of tools

Diffusion of innovations mechanism



**12 popular quality assurance tools**

Continuous integration	Dependency management	Code coverage reporters	Cross browser testers
build passing	dependencies up to date	coverage 94%	Firefox 82 Windows 7 ✓   Chrome 86 Windows 7 ✓
Travis Circle Appveyor Codeship	David Bithound Gemnasium	Coveralls Codeclimate Codecov Codacy	Saucelabs

**For each tool:**

**Heterogeneous network**

```
graph TD; R1((R1)) -- committer --> R2((R2)); R1 -- committer --> R3((R3)); R2 -- dependency similarity --> R7((R7)); R2 -- pull req --> R4((R4)); R3 -- watcher --> R1; R3 -- dependencies --> R6((R6)); R4 -- description similarity --> R5((R5)); R5 -- dependencies --> R6; R6 -- dependencies --> R4;
```

**Hazard modeling (Cox regression)**

A Kaplan-Meier survival plot showing survival probability over time (0 to 60). Two curves are shown: a blue curve for one group and an orange curve for another, both showing a decreasing trend over time.

# Today: Open source ecosystem as ...

---

- Highly **transparent** environment → Nudging mechanisms
- **Attention** economy → Supply and demand mismatch
- Socio-technical **network** → Network effects

