

Leveraging Signals to Build More Sustainable Open Source Communities

Bogdan Vasilescu
@b_vasilescu

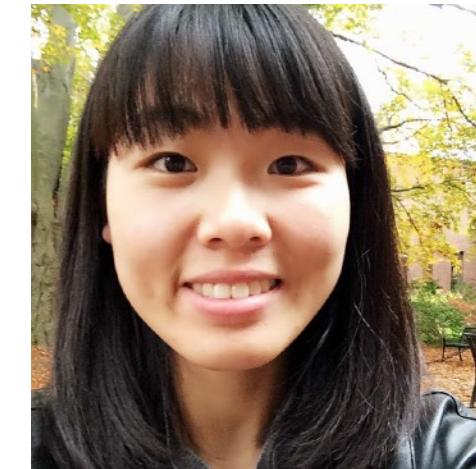
Acknowledgements



Courtney Miller



Anita Brown



Michelle Cao



Jim Herbsleb



Christian Kästner



David Widder



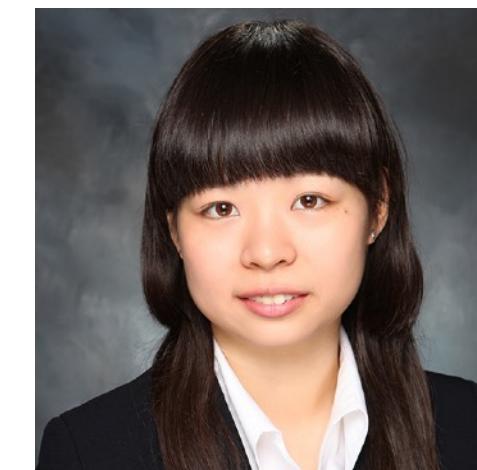
Anita Sarma



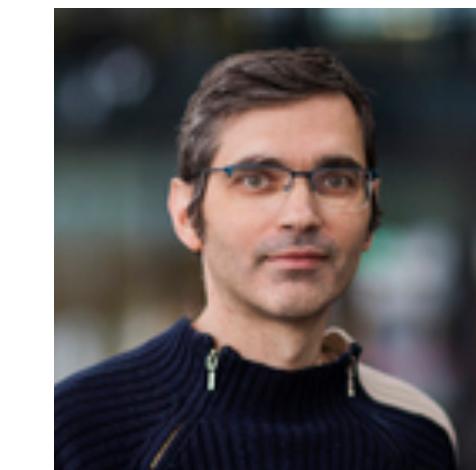
Audris Mockus



Alex Nolte



Sophie Qiu



Alex Serebrenik



Marat Valiev



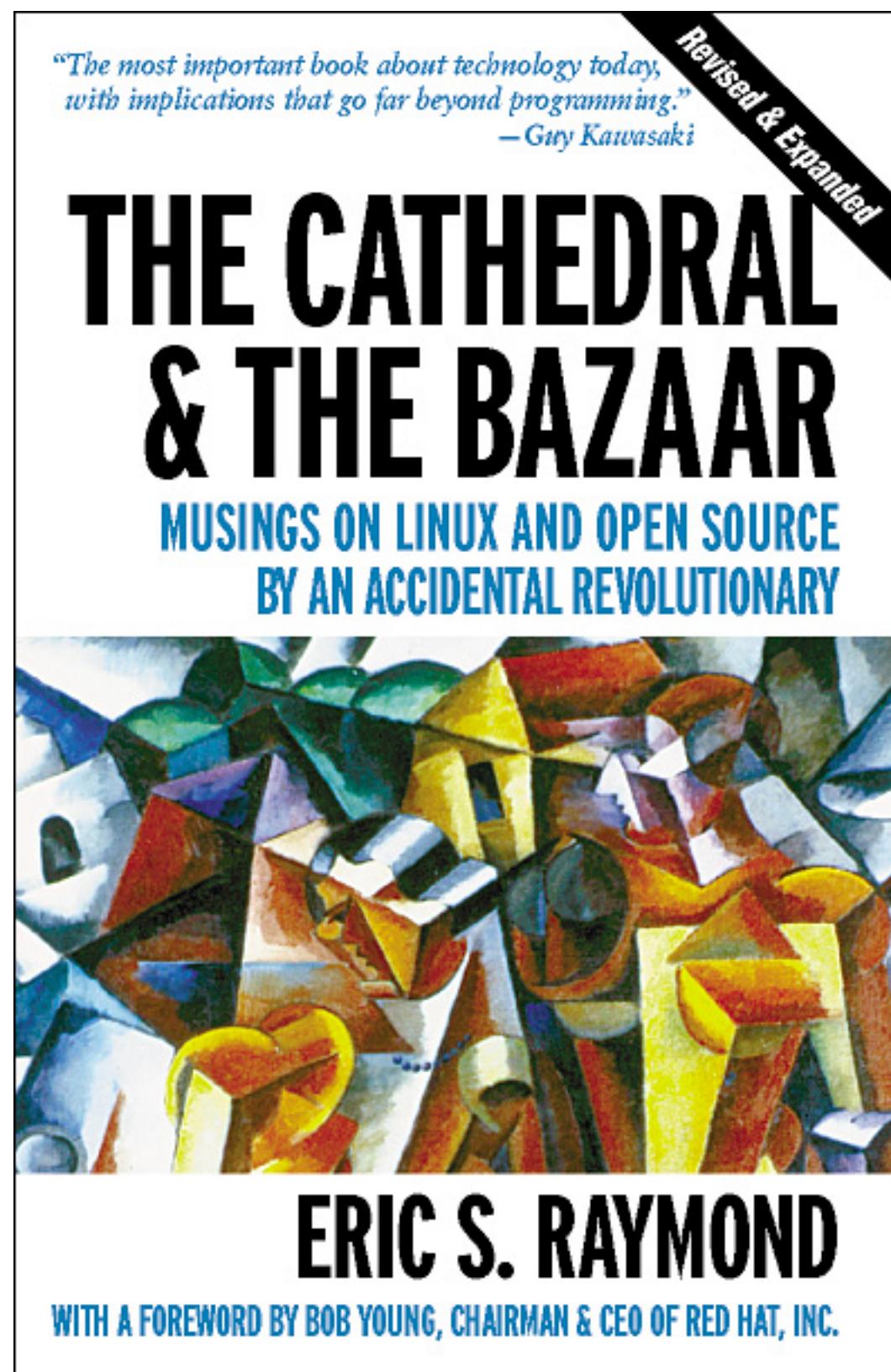
Laura Dabbish



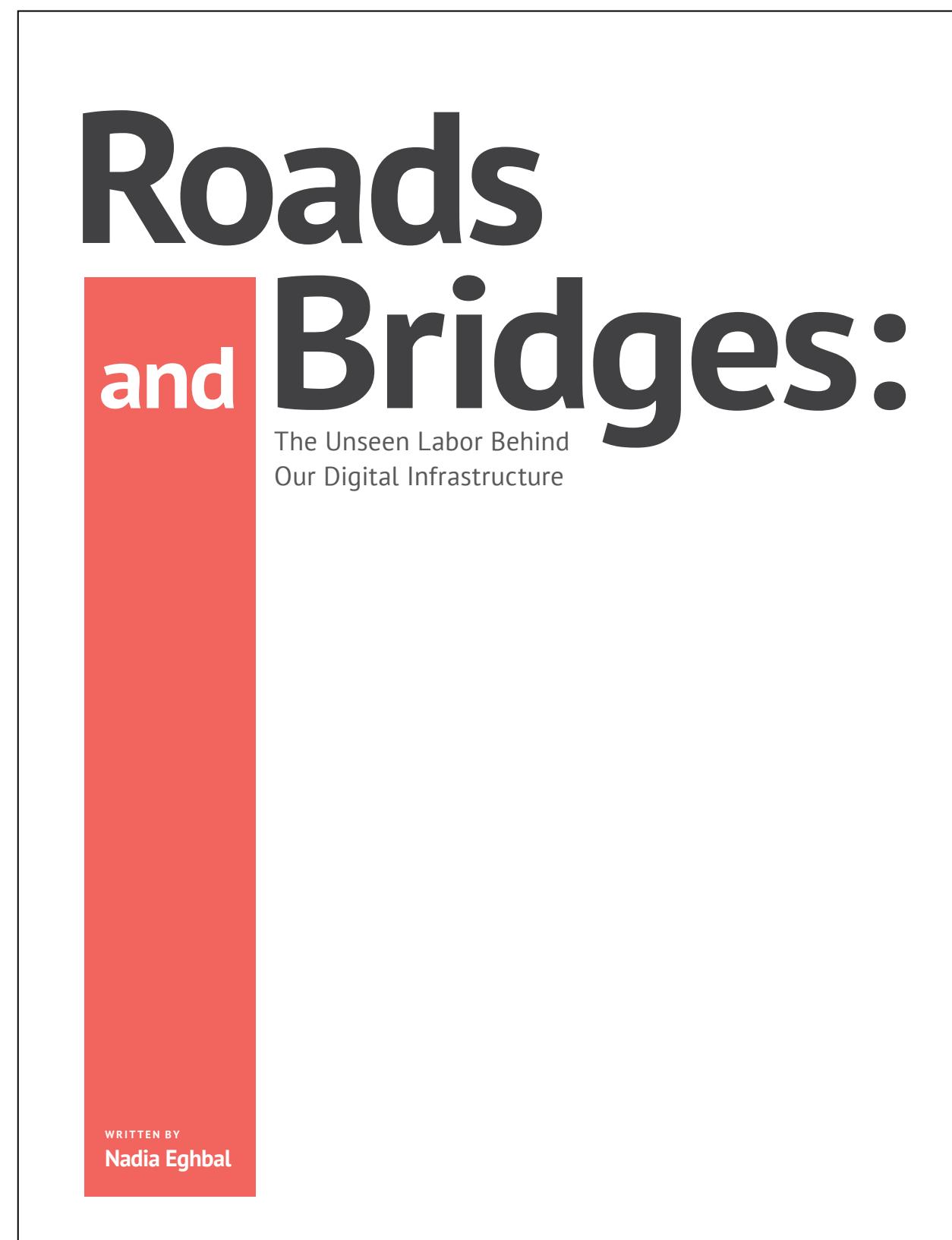
Lily Li

Open source software: from curiosity to digital infrastructure

1999



2016



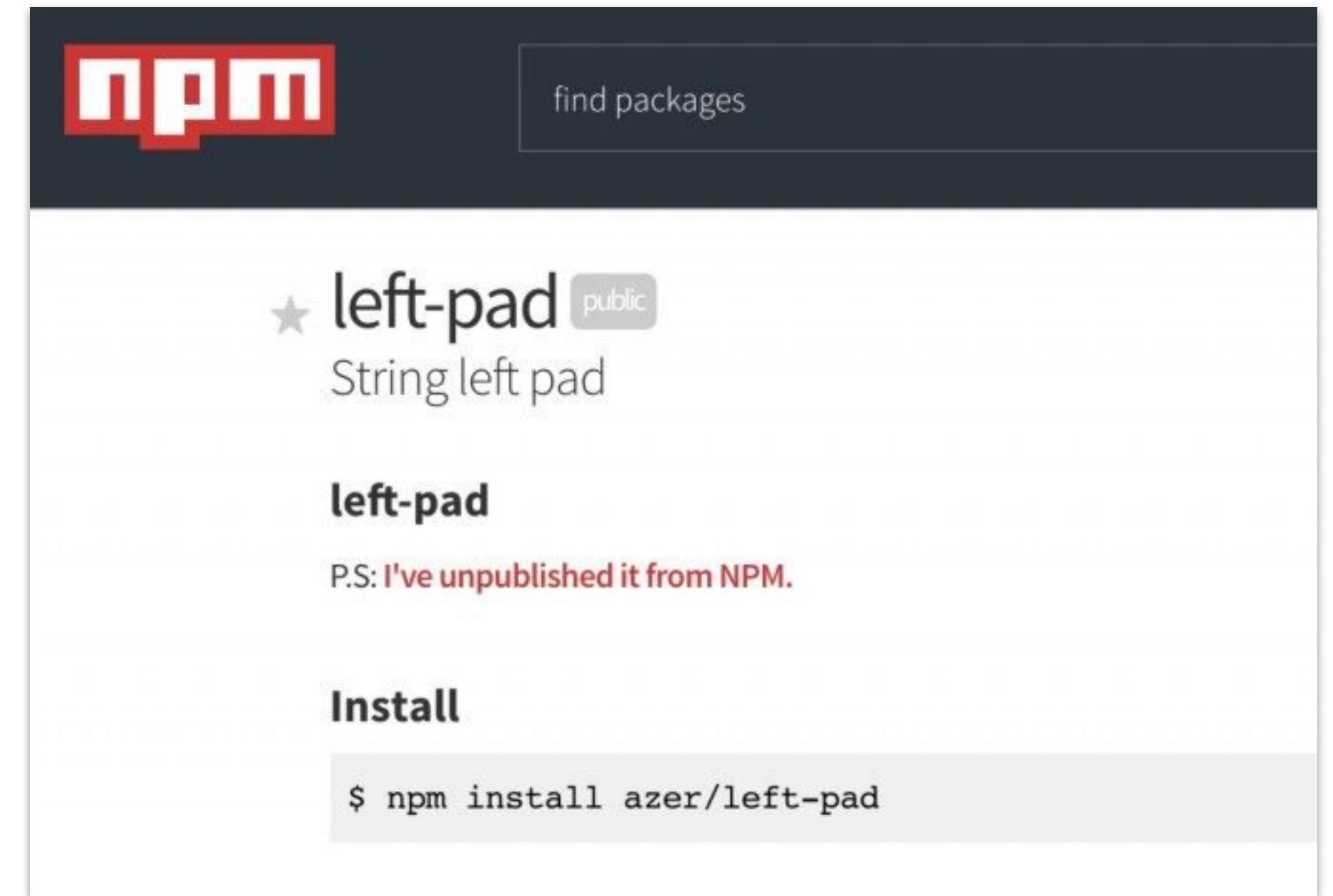
- Open source code as digital roads or bridges:
 - can be used by anyone to build software
- Nearly all software that powers our society relies on open source code
- Everybody uses open source code:
 - Fortune 500 companies
 - government
 - major software companies
 - startups

Economists: open source as “digital dark matter” i.e., important but mostly invisible

- Apache web server installations valued at \$7–\$10 billion in the US alone (Greenstein and Nagel, 2016)
- The economic value of open source software to Europe totaled ~456 billion Euros per year in 2010 (Daffara, 2012)
- There are millions of other open source projects besides the Apache web server, many in similarly important roles

Just like physical infrastructure, digital infrastructure needs regular upkeep and maintenance

- Risks for downstream users from depending on abandoned or unmaintained libraries
 - Security breaches, interruptions in service, ...
 - Leftpad
 - OpenSSL + Heartbleed
- Also slows down innovation
 - Startups rely heavily on this infrastructure



Open source needs a **steady supply of effort by contributors**

But that is **harder today than ever before**
... because of how open source has **changed**

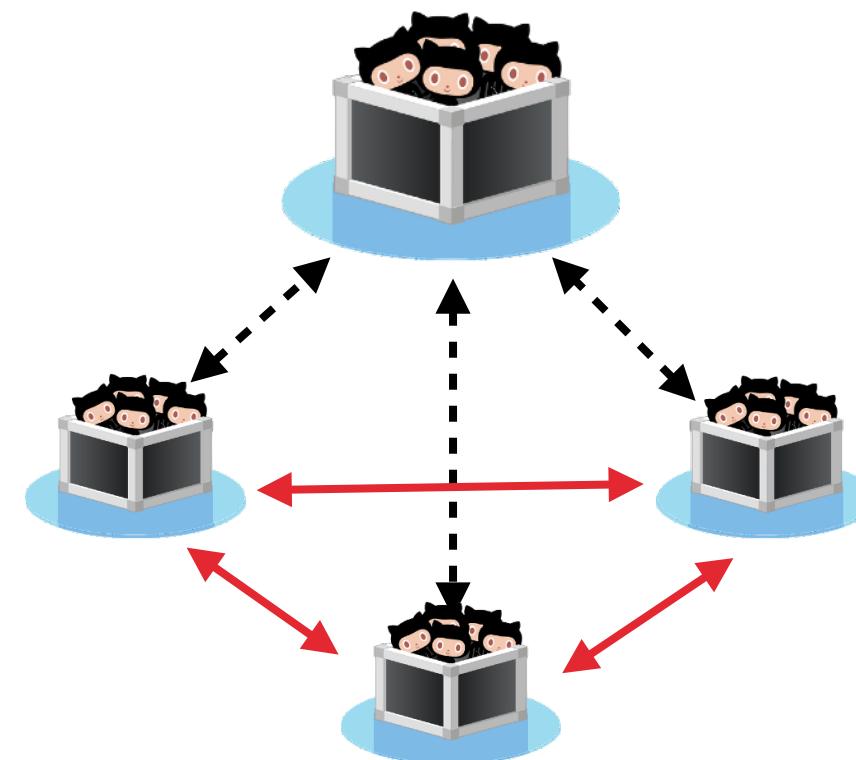


Today: more problems than solutions

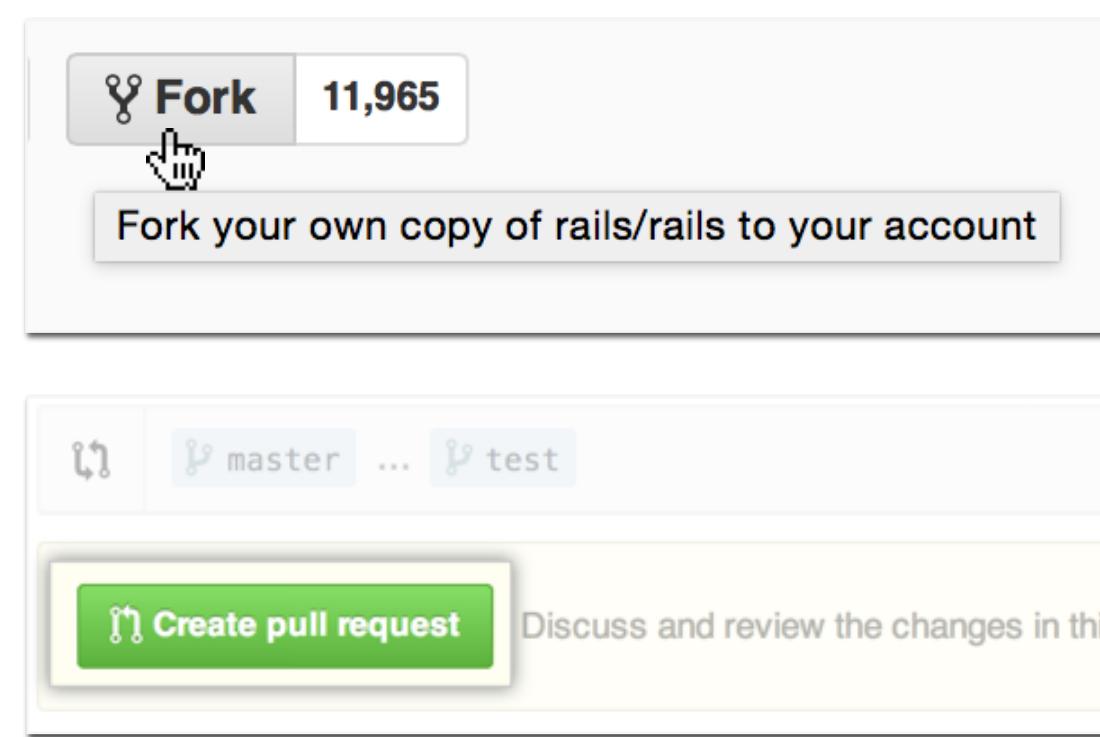
How has open source changed?

Change #1: GitHub as a standardized place to collaborate on code

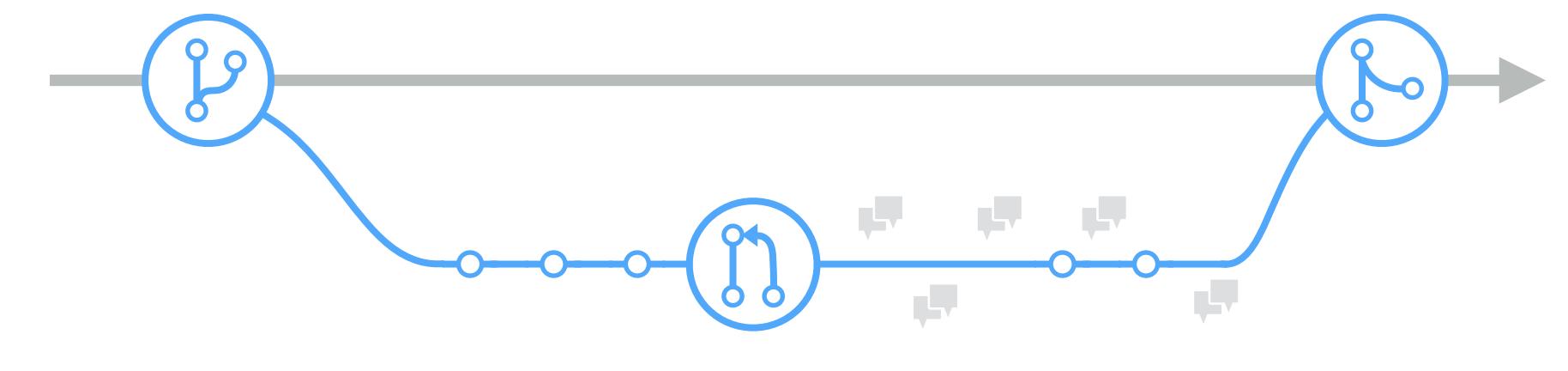
- Git version control



- GitHub UI



- The Pull Request model



- Lower barrier to entry
- Easier to contribute



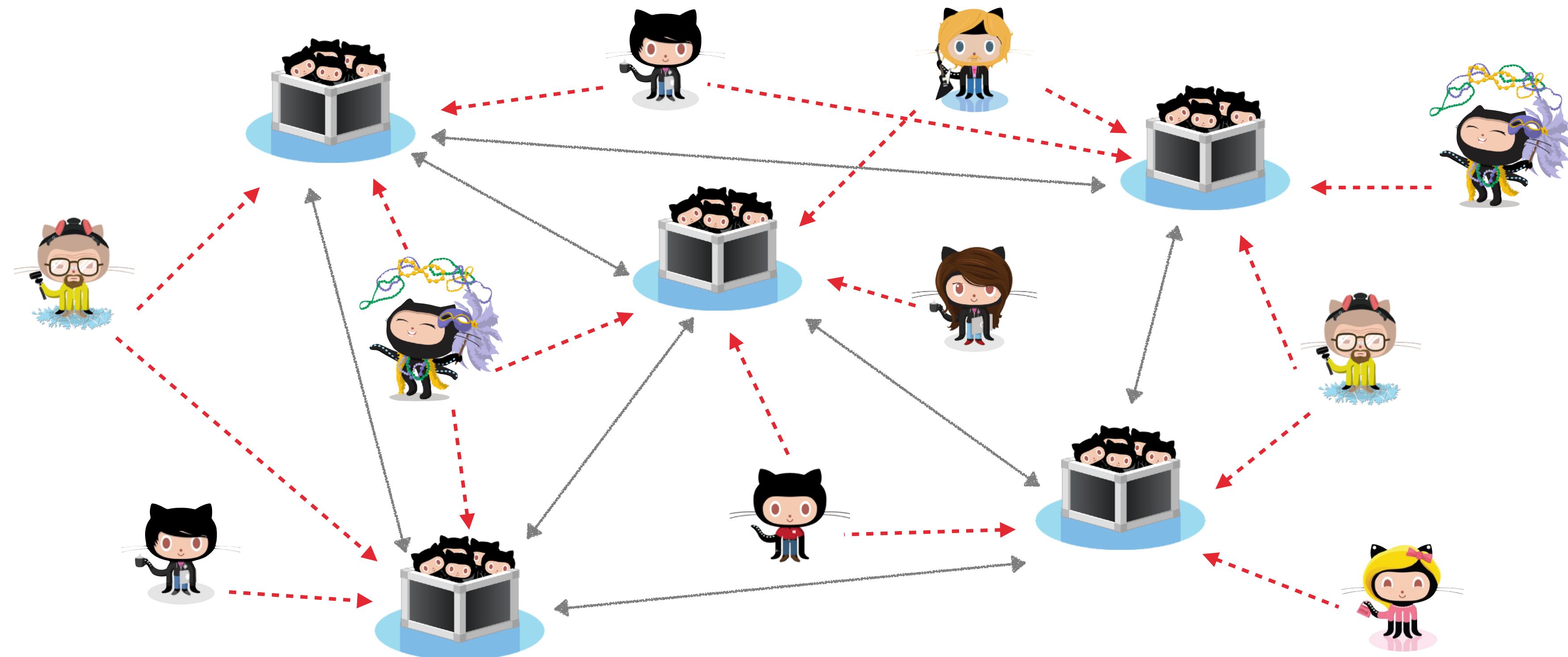
More production

More open source code now than ever before

- Explosion of production in the past seven years



Change #2: Complex *ecosystem* of interdependencies



Socio-technical environment: heterogeneous links

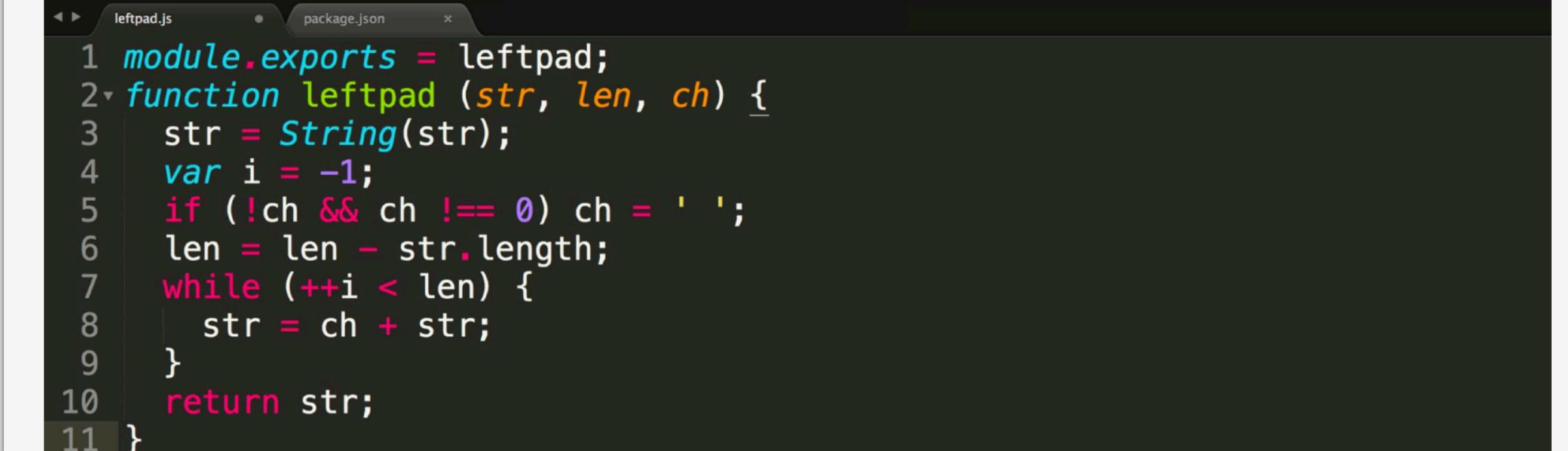
Network effects

- Leftpad-like incidents
- Breaking changes
 - ▶ (Bogart et al. 2016)
- Tangled issue reports
 - ▶ (Ma et al. 2017), (Zhang et al 2018)
- ...

NPM ERR!

How one programmer broke the internet by deleting a tiny piece of code

By Keith Collins • March 27, 2016



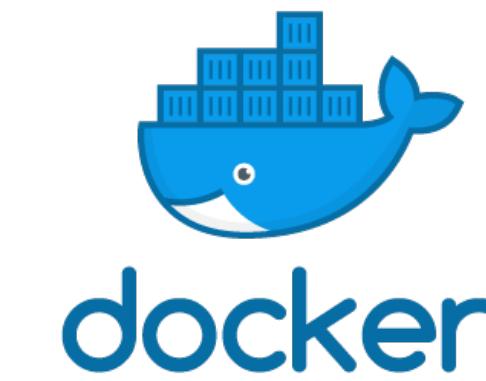
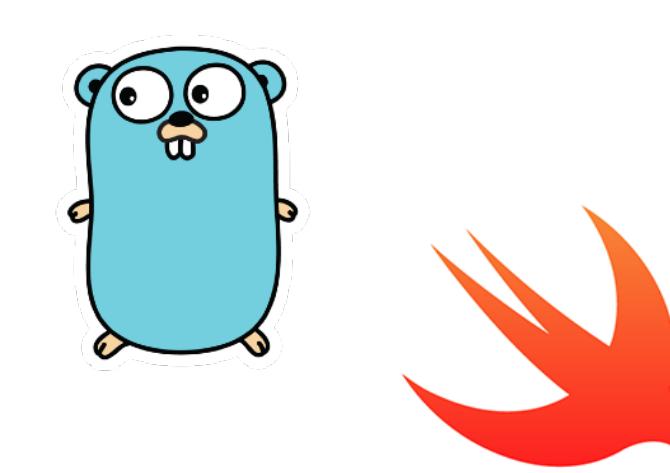
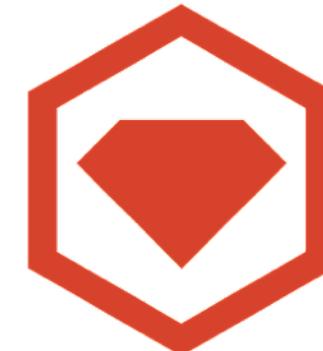
```
1 module.exports = leftpad;
2 function leftpad (str, len, ch) {
3     str = String(str);
4     var i = -1;
5     if (!ch && ch !== 0) ch = ' ';
6     len = len - str.length;
7     while (++i < len) {
8         str = ch + str;
9     }
10    return str;
11 }
```

<https://qz.com/646467/how-one-programmer-broke-the-internet-by-deleting-a-tiny-piece-of-code/>

- Within-Ecosystem Issue Linking: A Large-scale Study of Rails. Zhang, Y., Yu, Y., Wang, H., Vasilescu, B., and Filkov, V. *Software Mining Workshop 2018*

Change #3: Increasing commercialization and professionalization

- Historically
 - Community-based projects (Python, RubyGems, Twisted)
- Currently
 - Lots of commercial involvement
 - Companies (Go - Google, React - Facebook, Swift - Apple)
 - Startups (Docker, npm, Meteor)



- 23% of respondents to 2017 GitHub survey: job duties include contributing to open source

<http://opensourcesurvey.org/2017/>

Change #4: High level of transparency

- Profile pages for users and projects
- Rich inferences about people's expertise and level of commitment
- Impacts collaboration, but also recruiting and hiring
 - (Dabbish et al. 2012), (Marlow et al. 2013), (Marlow and Dabbish 2013)

The image shows a composite view of GitHub features. At the top, a GitHub profile card for a user named 'caolan' is displayed, featuring a cartoon GitHub logo holding a CV. Below this, a list of 'Popular repositories' is shown, including 'breakfast-repo', 'x86-kernel', and 'jsconf-2015-deck'. To the right, a sidebar lists 'Repositories contributed to' such as 'npm/docs', 'mozilla/publish.webmaker.org', and 'npm/marky-markdown'. The bottom half of the image shows a detailed view of the 'caolan / async' repository page. This page includes tabs for Code, Issues (21), Pull requests (6), Projects (0), Wiki, and Insights. It displays statistics like 1,629 commits, 11 branches, 72 releases, and 206 contributors. A large image of the 'async' logo is prominently featured. Below the logo, build status (passing), npm version (v2.6.0), coverage (99%), and other repository metrics are listed. A descriptive text block at the bottom explains that Async is a utility module for asynchronous JavaScript.

How have these changes affected the open source communities?

High expectations toward the quality, reliability, and security of open source infrastructure

- Equifax (market cap \$14 billion) built products on top of open-source infrastructure, including Apache Struts
- Equifax did not make any contributions to open source projects
- A flaw in Apache Struts contributed to the breach (CVE-2017-5638)
- Equifax publicly blamed (with national news coverage) Apache Struts for the breach

Equifax confirms Apache Struts security flaw it failed to patch is to blame for hack

The company said the March vulnerability was exploited by hackers.



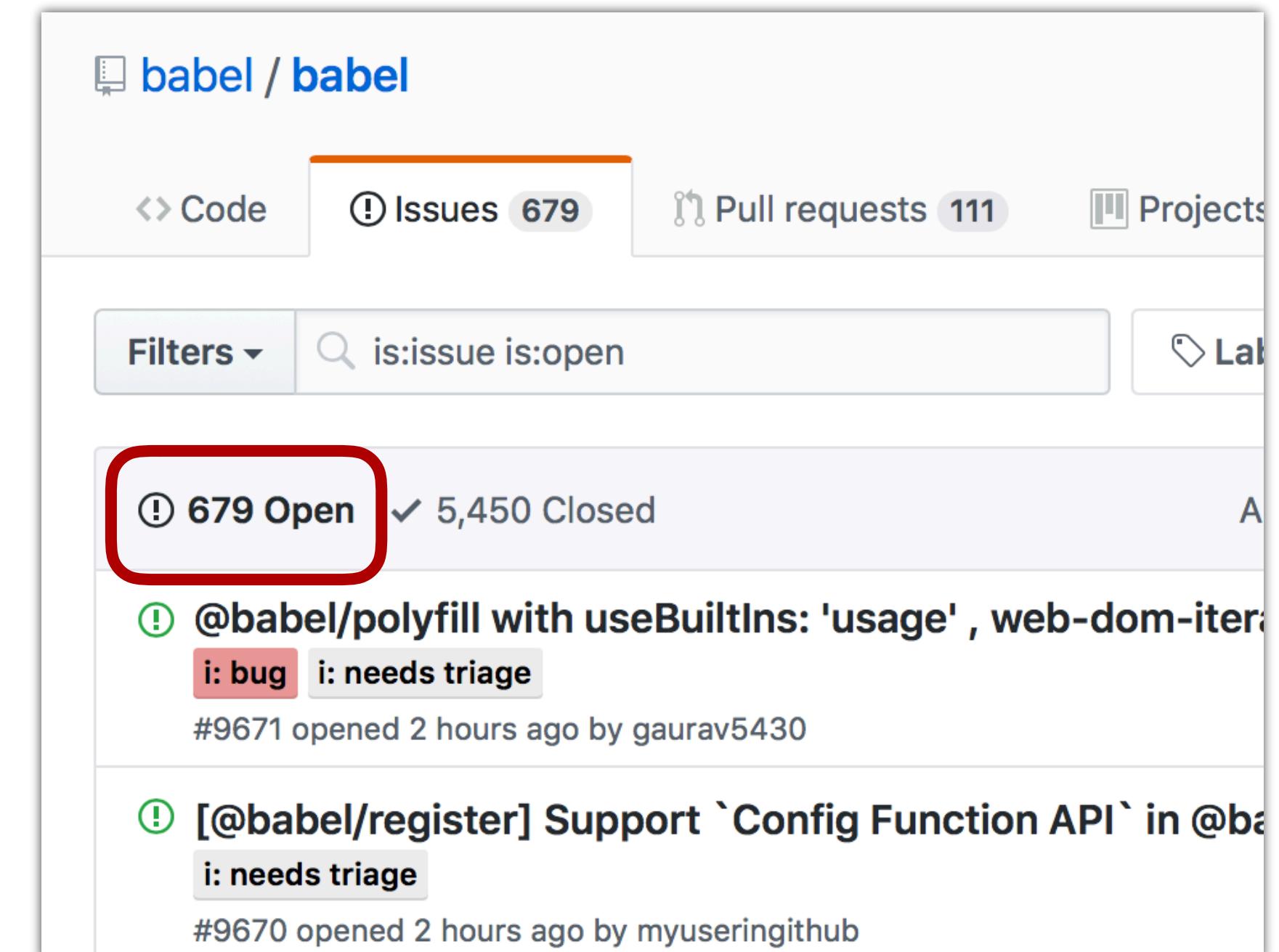
By Zack Whittaker | September 14, 2017 -- 01:27 GMT (18:27 PDT) | Topic: Security



<https://www.zdnet.com/article/equifax-confirms-apache-struts-flaw-it-failed-to-patch-was-to-blame-for-data-breach/>

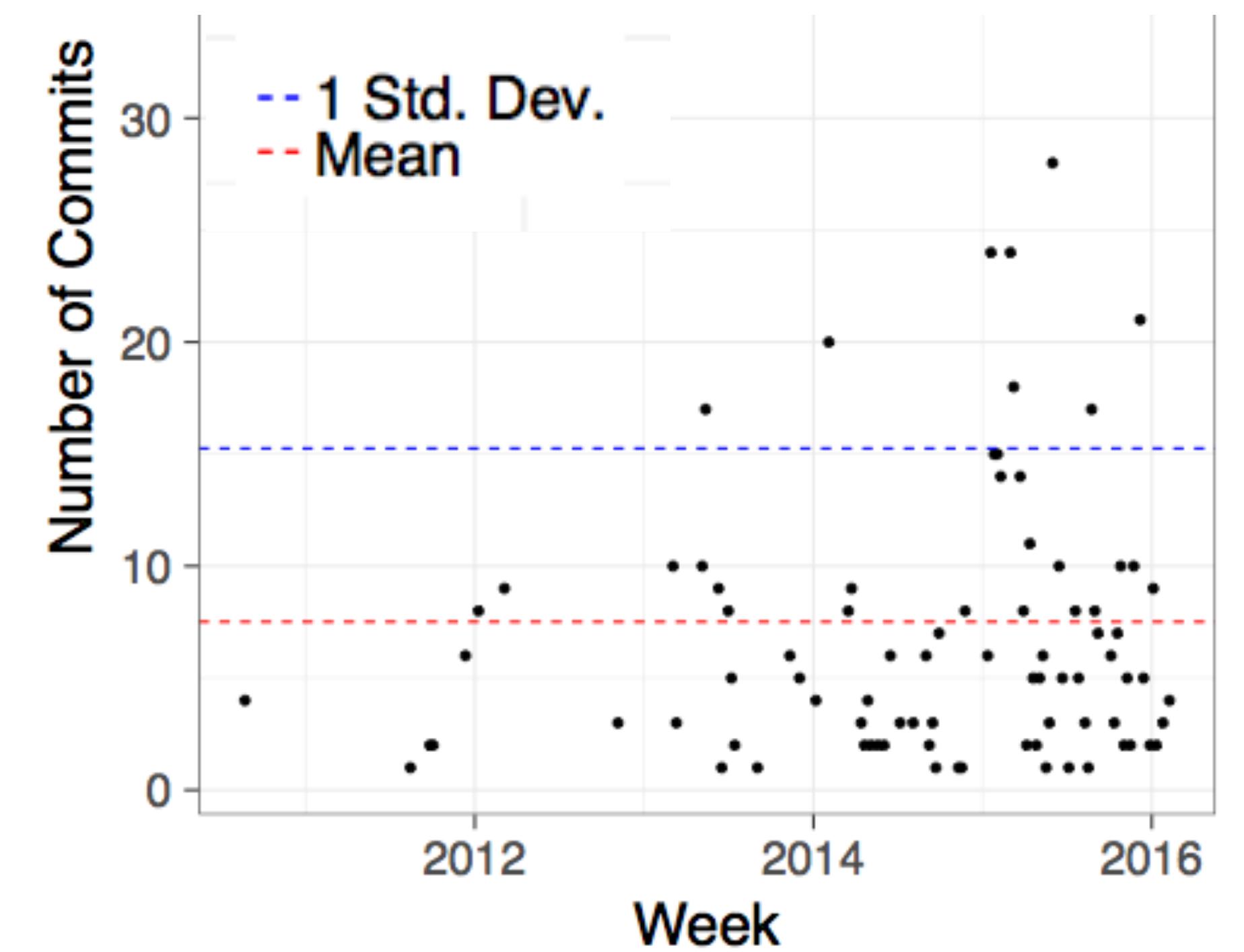
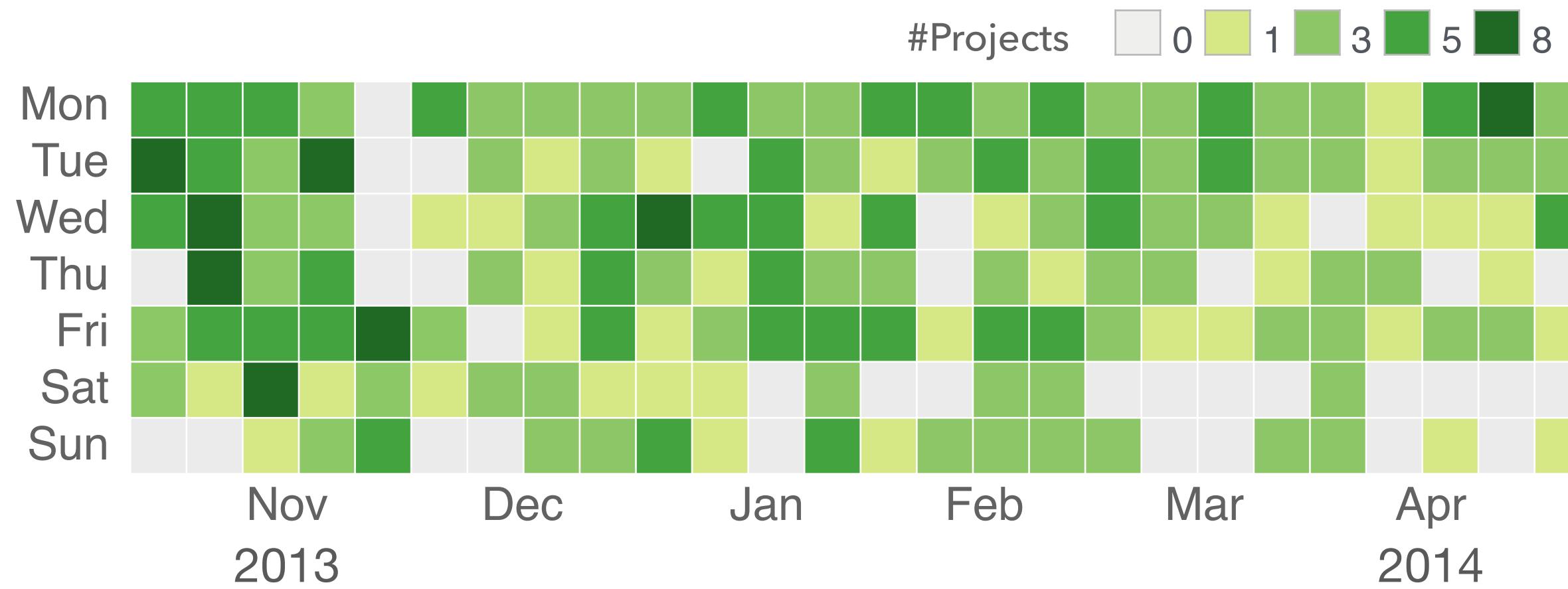
High level of demands & stress

- Easy to report issues / submit PRs
 - Growing volume of requests
- Social pressure to respond quickly
 - Otherwise, off-putting to newcomers (Steinmacher et al. 2015)
- Entitlement, unreasonable requests from users:
 - “*I have been waiting 2 years for Angular to track the ‘progress’ event and it still can’t get it right?!?!*”
 - “*Thank you for your ever useless explanations.*”



High-workload, potentially high-stress environment

- Working on many projects concurrently
- Higher than average workload



- The Sky is Not the Limit: Multitasking on GitHub Projects. Vasilescu, B., Blincoe, K., Xuan, Q., Casalnuovo, C., Damian, D., Devanbu, P., and Filkov, V. *ICSE 2016*

- Socio-Technical Work-Rate Increase Associates With Changes in Work Patterns in Online Projects. Sarker, F., Vasilescu, B., Blincoe, K., and Filkov, V. *ICSE 2019*

Example: “Longest streak” backlash

Contribution graph can be harmful to contributors #627

! Open

mxsasha opened this issue on Apr 1, 2016 · 189 comments



mxsasha commented on Apr 1, 2016

A common well-being issue in open-source communities is the tendency of people to over-commit. Many contributors care deeply, at the risk of saying yes too often harming their well-being. Open-source communities are especially at risk, because many contributors work next to a full-time job.

...

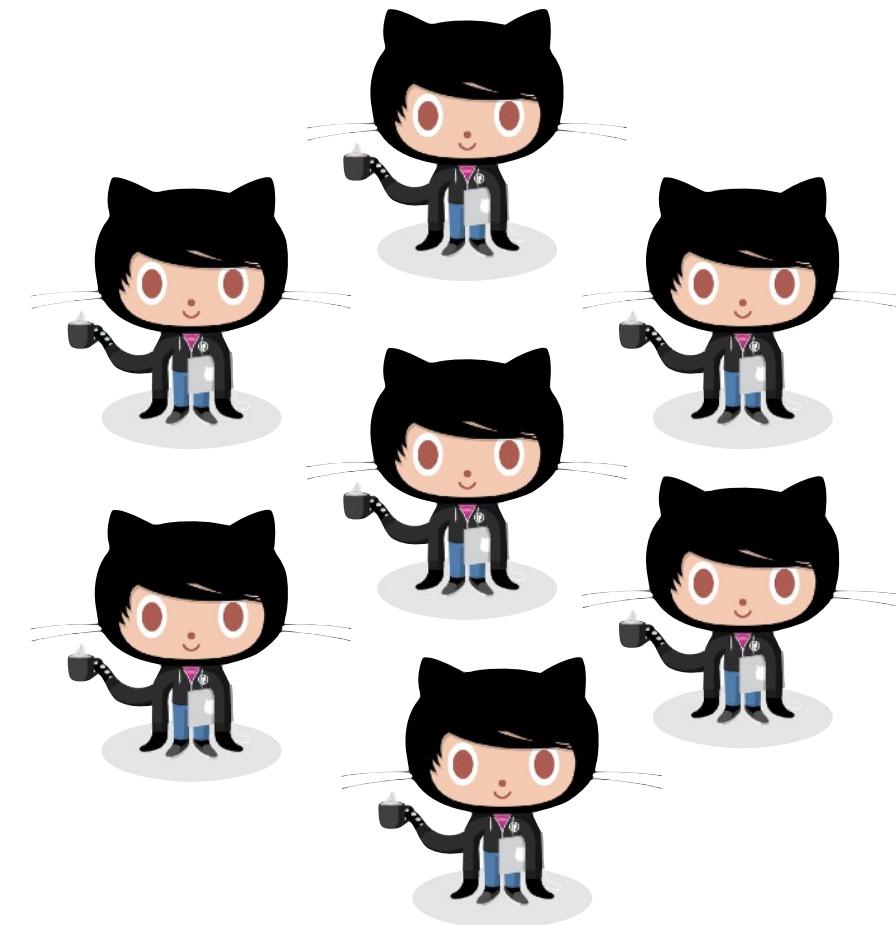
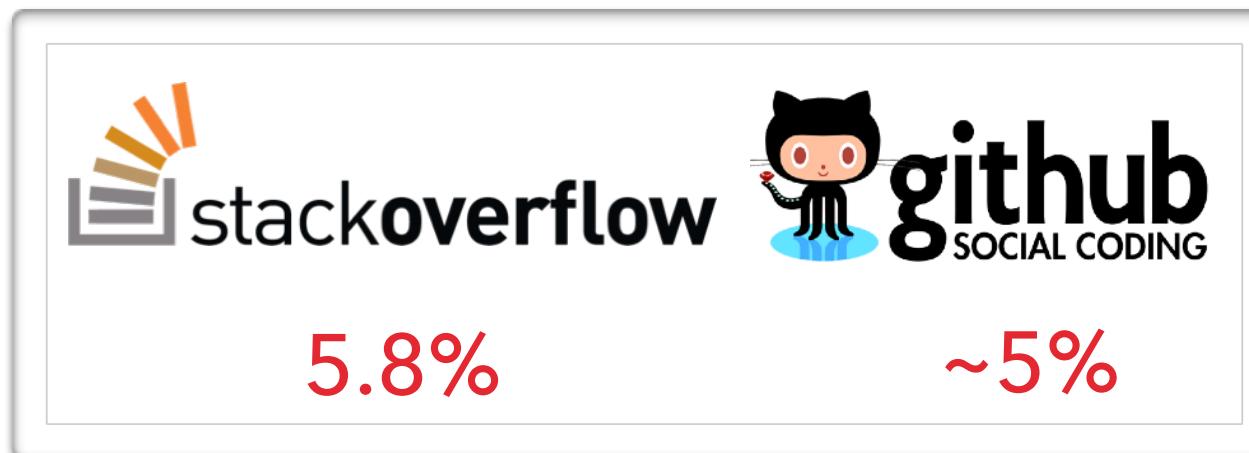
Any mechanism in our community that motivates people to avoid taking breaks and avoid stepping back, can be harmful to the well-being of contributors and is thereby harmful to open source as a whole. Even though it was probably introduced with the best intentions. If our interests are really in supporting open-source long-term, this graph should be removed or substantially changed so that it no longer punishes healthy behaviour. For example, what if we would give people achievements for taking breaks instead of working non-stop?

I therefore want to ask you to consider removing or substantially changing the contribution graph and its related statistics, to help guard the well-being of the contributors and the communities.

I also wrote about this in a bit more detail on my blog: <http://erik.io/blog/2016/04/01/how-github-contribution-graph-is-harmful/>

Low demographic diversity

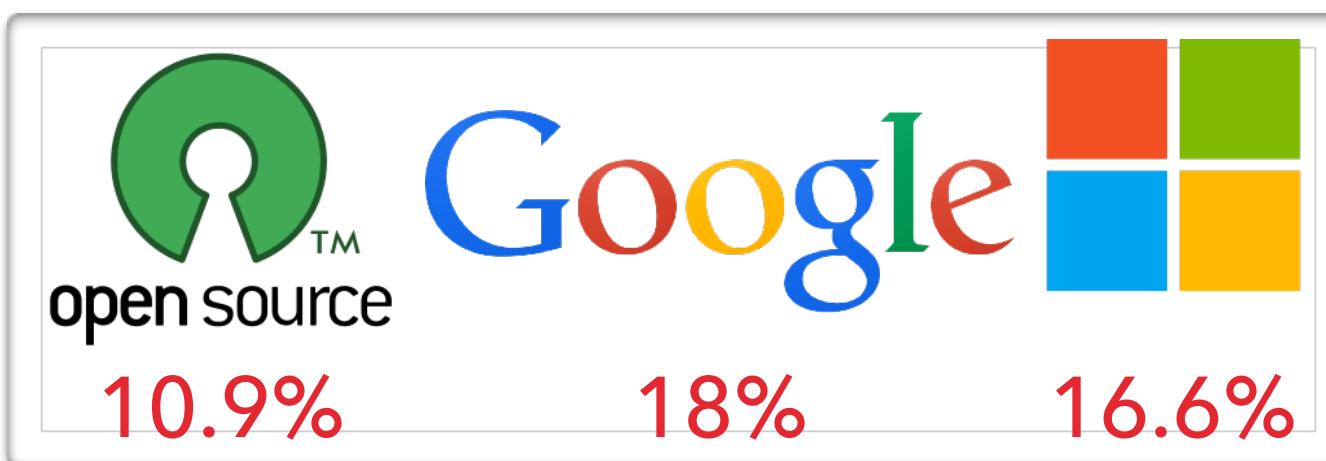
- Gender representation reality



- Expectation



“More about the contributions to the code than the ‘characteristics’ of the person”



“Any demographic identity is irrelevant”

“Code sees no color or gender”

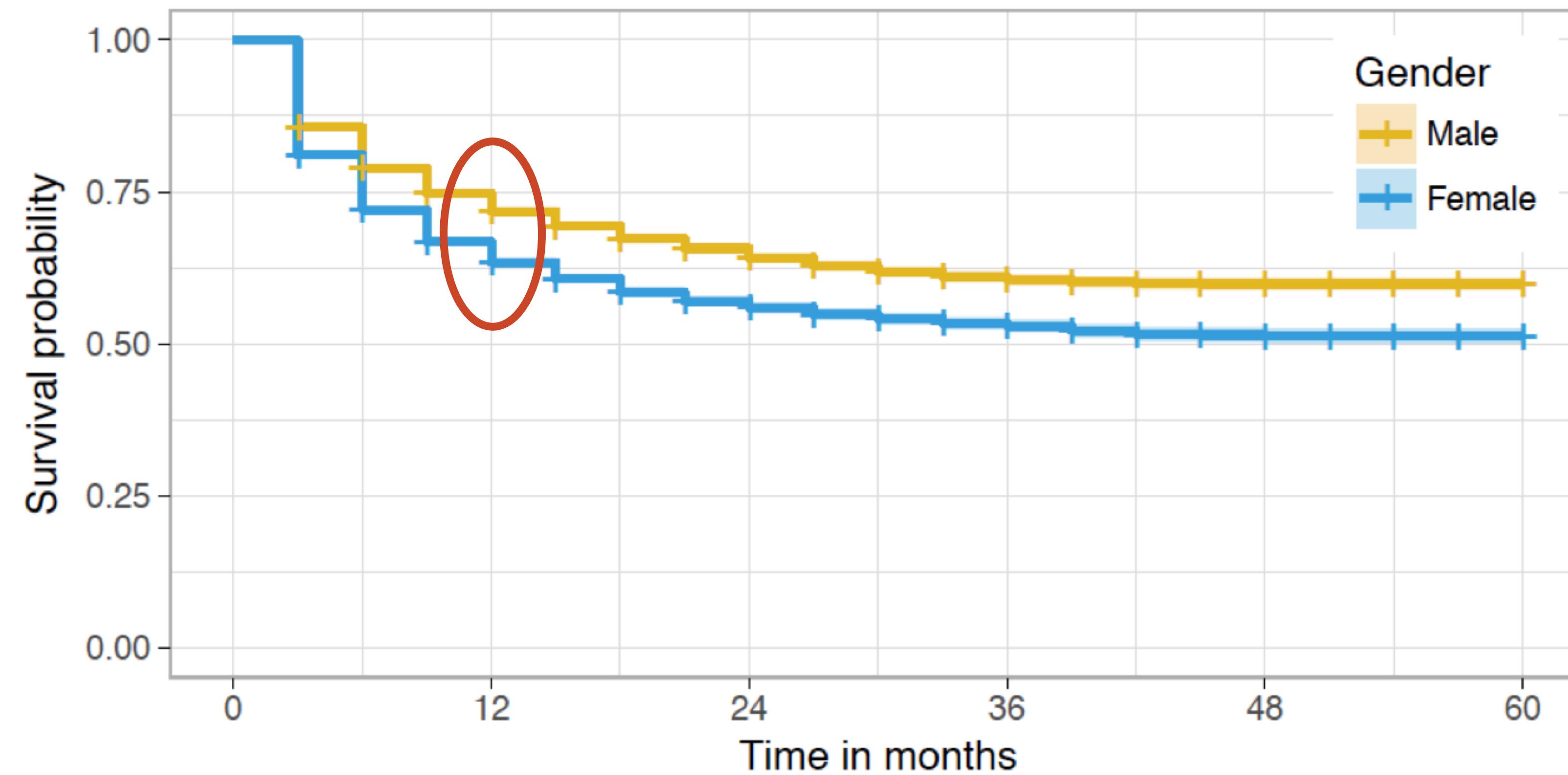
- FLOSS 2013: A survey dataset about free software contributors: challenges for curating, sharing, and combining G Robles, L Arjona-Reina, B Vasilescu, A Serebrenik, JM Gonzalez-Barahona. MSR 2014
- Google Diversity (2015) www.google.com/diversity/index.html#chart
- Inside Microsoft (2015) <https://goo.gl/nT4YiI>

- Exploring the data on gender and GitHub repo ownership Alyssa Frazee. <http://alyssafrazee.com/gender-and-github-code.html>
- Stack Overflow 2015 Developer Survey (26,086 people from 157 countries) <http://stackoverflow.com/research/developer-survey-2015#profile-gender>

- Perceptions of Diversity on GitHub: A User Survey. Vasilescu, B., Filkov, V., and Serebrenik, A. CHASE 2015

On GitHub, women disengage earlier than men

After one year ~70% of men are still active but only ~60% of women*



*Among committers with first & last names on their profiles

• Going Farther Together: The Impact of Social Capital on Sustained Participation in Open Source.
Qiu, H.S., Nolte, A., Brown, A., Serebrenik, A., and Vasilescu, B. ICSE 2019

What are people doing to
attract & retain (diverse)
contributors to open source?

Recent trend: Guides for newcomers

The image shows two side-by-side screenshots. On the left, a screenshot of a GitHub repository page titled 'GitHub Help' shows a 'FIRST TIMERS ONLY' guide. The guide features a large red banner at the top with the text 'FIRST TIMERS ONLY' and 'Friendly Open Source projects should be the first thing you look at'. Below this, there's a section titled 'Welcome! Let's get started' with text about contributing to open source. On the right, a screenshot of the CodeTriage website has a dark blue background. It features a logo with four colored squares (red, green, blue, yellow) followed by the text 'CodeTriage'. Below this, a large heading reads 'Help out your favorite open source projects and become a better developer while doing it.' A paragraph explains that users can pick repos to receive daily issues and fix them, with 41,383 developers working on 4,275 repos. A 'SIGN UP WITH GITHUB' button is at the bottom.

GitHub Help GitHub...

FIRST TIMERS ONLY
Friendly Open Source projects should be the first thing you look at

About Log in Sign Up

Helping new contributors find a project with labels

Apply the `help wanted` issue labels to issues in your repository to highlight them for people to contribute to.

Welcome! Let's get started

Contributing to open source can be a little overwhelming. Perhaps you're maybe you've been coding for a while, but you're not sure which project you felt comfortable contributing to.

You can do

If you have never contributed to an open source project and you're just getting started, here are some resources.

CodeTriage

Help out your favorite open source projects and become a better developer while doing it.

Pick your favorite repos to receive a different open issue in your inbox every day. Fix the issue and everybody wins. 41,383 developers are working on 4,275 open source repos using CodeTriage.

What is CodeTriage?

SIGN UP WITH GITHUB

Recent trend: Codes of conduct

The screenshot shows a section of the GitHub Help website titled "Adding a code of conduct to your project". It includes a large heading, a paragraph explaining the purpose of a code of conduct, and a smaller paragraph defining it. A search bar is visible at the top right.

Adding a code of conduct to your project

Adopt a code of conduct to define community standards, signal a welcoming and inclusive project, and outline procedures for handling abuse.

A *code of conduct* defines standards for how to engage in a community. It signals an inclusive environment that respects all contributions. It also outlines procedures for addressing problems

The screenshot shows the homepage of the Contributor Covenant. It features a purple header with the logo and name, followed by a main content area with a white background. The content discusses the importance of diversity in open source projects and the role of codes of conduct in creating welcoming environments.

A Code of Conduct for Open Source Projects

Open Source has always been a foundation of the Internet, and with the advent of social open source networks this is more true than ever. But free, libre, and open source projects suffer from a startling lack of diversity, with dramatically low representation by women, people of color, and other marginalized populations.

Often it is the unintentional assumptions and actions of project maintainers and participants that make open source projects unwelcoming (or even hostile) to marginalized people: making assumptions about gender or race, reinforcing stereotypes, using sexualized or otherwise inappropriate language, or demonstrating a lack of regard for the safety and well-being of vulnerable people.

One way to begin addressing this problem is to be overt in our openness, welcoming all people to contribute, and pledging in return to value them as whole human beings and to foster an atmosphere of kindness, cooperation, and understanding.

Adopting the Contributor Covenant can be one way to express and codify these values and signal your intention to make your open source community welcoming, diverse, and inclusive.

(Tourani, Adams, & Serebrenik, SANER 2017)

Recent trend: Safe spaces

Table 1. OSS websites with women only spaces.

| Software package | Name of the space | URL |
|------------------|---|---|
| ArchLinux | Arch Linux for Women | http://archwomen.org |
| Bitcoin | Women in Bitcoin Madchenabend in Berlin | https://www.facebook.com/womeninbitcoin/ |
| BonitaSoft | Blog Post about Community efforts for encouraging women | https://community.bonitasoft.com/behind-scenes-bonita-21-27-feb-2011 |
| Debian | Debian Women | https://www.debian.org/women |
| Drupal | Women in Drupal | http://www.womenindrupal.org/ |
| Fedora | Fedora Women | http://fedoraproject.org/wiki/Women |
| FreeNX | IRC Channel for Women | https://archwomen.org/wiki/aw-org:irc |
| GNOME | GNOME Women | https://wiki.gnome.org/GnomeWomen http://gnome.org/opw/ |
| KDE | IRC Channel for Women | https://userbase.kde.org/IRC_Channels |
| Mozilla | WoMoz | http://www.womoz.org/blog/ |
| PHP | PHP Women | http://phpwomen.org/ |
| Ubuntu | Ubuntu Women Project | https://wiki.ubuntu-women.org/ |

(Singh & Brandon, OSS 2019)

Recent trend: Summer of Code



2018: “11.63% of accepted students are women”

A screenshot of the Rails Girls Summer of Code website. The header features the "Rails Girls Summer of Code" logo with a heart icon. On the right side of the header are links for "Blog", "Donate", and a menu icon. The main content area has a dark background with a photo of people working on laptops. It features the "1M Women to Tech" logo, which includes a stylized pink swirl and glasses. Below the logo is the tagline "FIND THE SKILLS AND OPPORTUNITIES YOU HAVE BEEN LOOKING FOR". There are three callout boxes: one with a location pin icon and "Online, Worldwide"; one with a calendar icon and "July 14 - October 14"; and one with a diamond icon and "19000 Women and still counting". A large button labeled "JOIN FREE" is prominent. At the bottom, there are social media icons for Facebook, YouTube, Twitter, Instagram, LinkedIn, and GitHub.

<https://opensource.googleblog.com/2018/06/google-summer-of-code-2018-statistics-part-2.html>

Recent trend: Hackathons

The screenshot shows the ANITA B.ORG website with a navigation bar featuring links for "WHAT WE DO", "WHO WE ARE", "EVENTS", and "OUR COMMUNITY". Below the navigation is a banner for the "Grace Hopper Celebration" with options to "Register", "Awards", "Speakers", "Sponsors", and "Acade...". The main content area displays information about the "2018 Open Source Day" on Thursday, September 27, from 9:30 a.m. to 5:30 p.m., describing it as a day-long hackathon for women.

ANITA B.ORG

WHAT WE DO ▾ WHO WE ARE ▾ EVENTS ▾ OUR COMMUNITY ▾

Grace Hopper Celebration Register Awards Speakers Sponsors Academic...

2018 Open Source Day

Thursday, September 27 | 9:30 a.m. – 5:30 p.m.

Open Source Day (OSD) provides women of all skill levels and backgrounds with the opportunity to collaborate and make a change. College students, professionals, experienced coders, and beginners join this day-long hackathon to develop projects for improving the world we live in.

PEARL HACKS

Feb. 16-17, 2019
UNC-Chapel Hill

REGISTER NOW

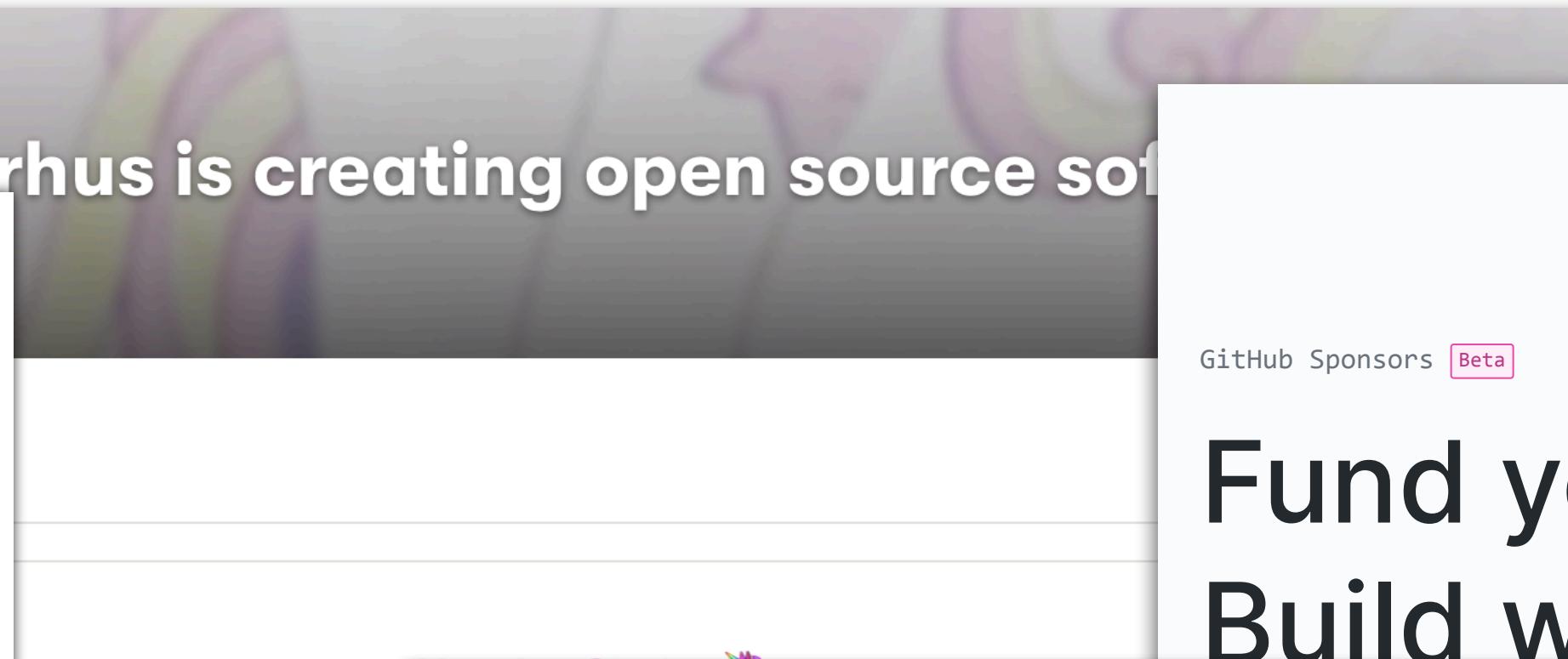
Recent trend: New forms of funding



Tidelift, a startup that aims to “make open source software work better for developers,” raised \$25 million in a series B round co-led by General Catalyst, Foundry Group, and former Red Hat Chairman and CEO Matthew Szulik.

The startup emerged out of stealth in May 2018 with backing from [General Catalyst](#) that led a \$15M Series A round in the company. Tidelift’s CEO and co-founder, Donald Fischer was also a venture partner at General Catalyst. Its other co-founders include Havoc Pennington, a former senior software architect at Continuum Analytics; Jeremy Katz, a former staff software engineer at Facebook.

<https://www.americaninno.com/boston/bostinno-bytes/open-source-software-marketplace-tidelift-raises-25m-in-series-b/>



I love open source. For many years I've been working on open source full-time, 8-12 hours a day while living off savings. I'm using Patreon now as my savings are slowly running out and I need help to continue my open source efforts.

I actively maintain [1100+ packages](#) ([2 billion downloads per month](#)) and [many popular projects](#). You're probably depending on some of my packages in your dependency tree. For example, [Webpack relies on 101 of my packages](#) and [Babel relies on 100 of my packages](#).

If you or your company are interested in supporting me, please consider backing me on [GitHub Sponsors](#) or [Patreon](#) to help me continue evolving these projects.

[See all my amazing supporters](#)

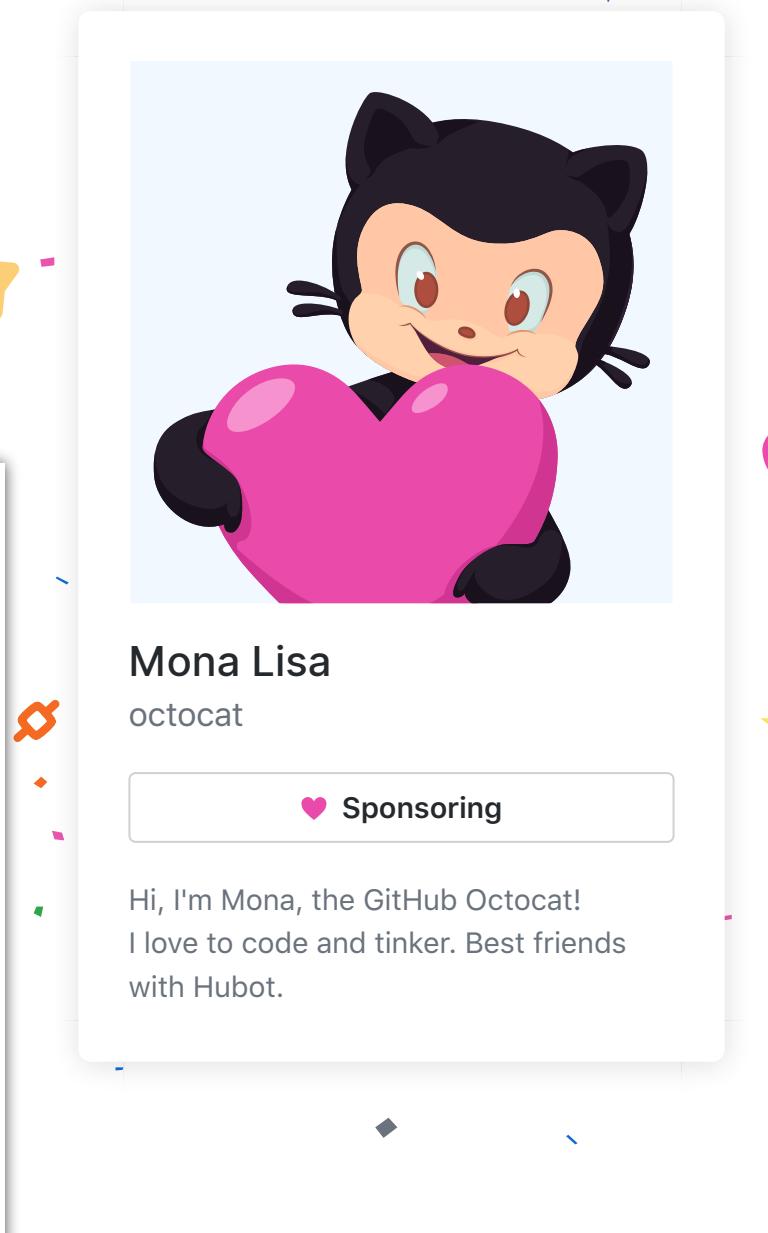
The program with a prize pool of almost US\$1 million aims to leverage the ‘power of the crowd’ in order to prevent another Heartbleed.

Tomáš Foltýn 7 Jan 2019 - 04:16PM

Share [Facebook](#) [Twitter](#) [LinkedIn](#)

The European Union (EU) is rolling out a bug bounty scheme on some of the most popular free and open source software around in a bid to ultimately make the internet a safer place.

A total of €851,000 (not too far from US\$1 million) is up for grabs as rewards for identifying security vulnerabilities in 15 widely used software projects (a full breakdown is shown below). A portion of the cash-for-bugs scheme is kicking off today, while nearly all others are scheduled to begin later this month.



<https://github.com/sponsors>

In summary:
Many possible interventions

Missing: THEORY

- When and where to apply which intervention?
- What effects to expect?
- What are the mediators / moderators?

But:
Huge potential for empirical research

The rest of this talk:
A few theory fragments

Example #1: It takes a village

Which projects are at risk of becoming abandoned?

Data:



70K PyPI packages

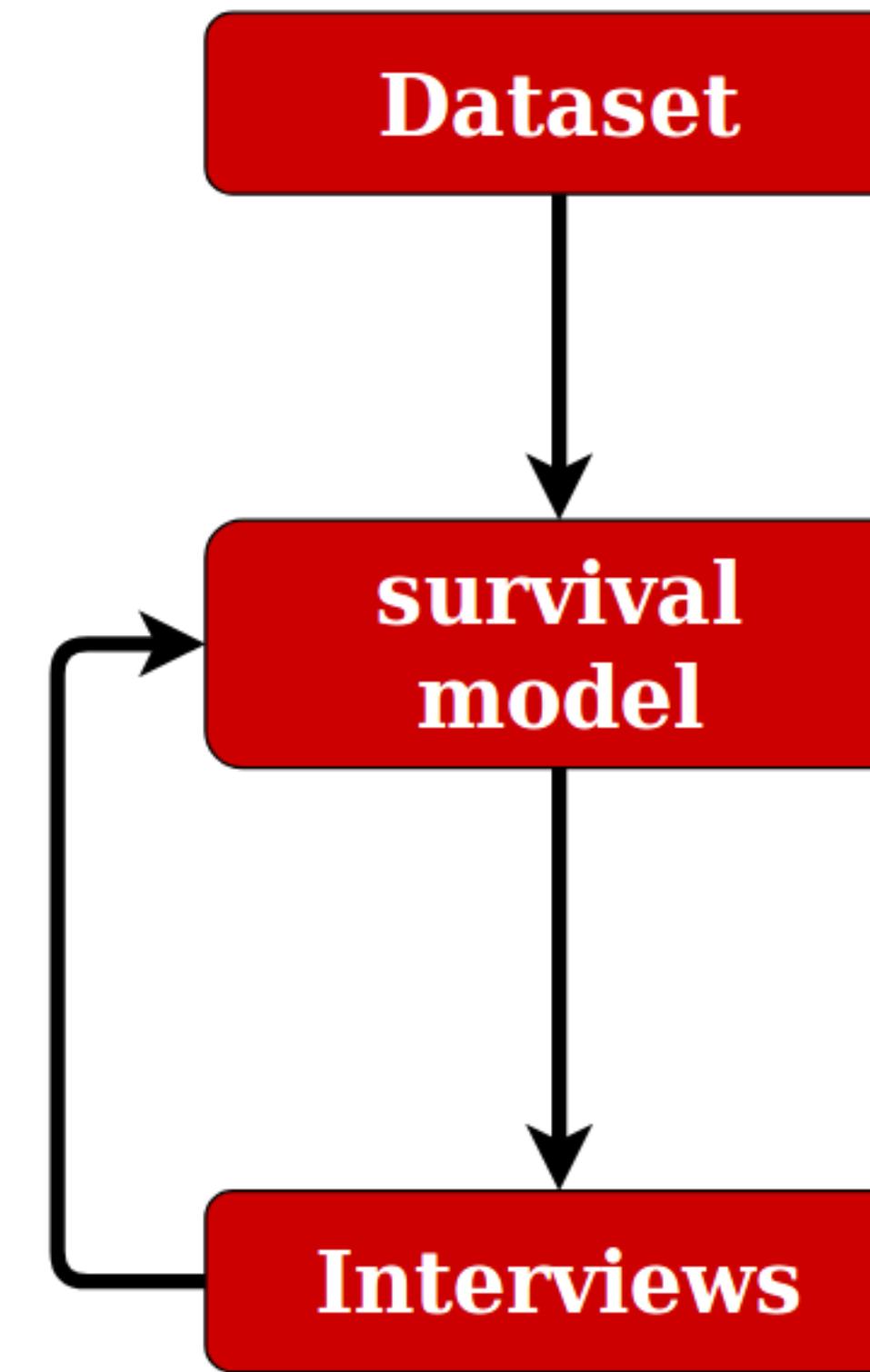
<https://zenodo.org/record/1297925>

Model:

Cox survival regression

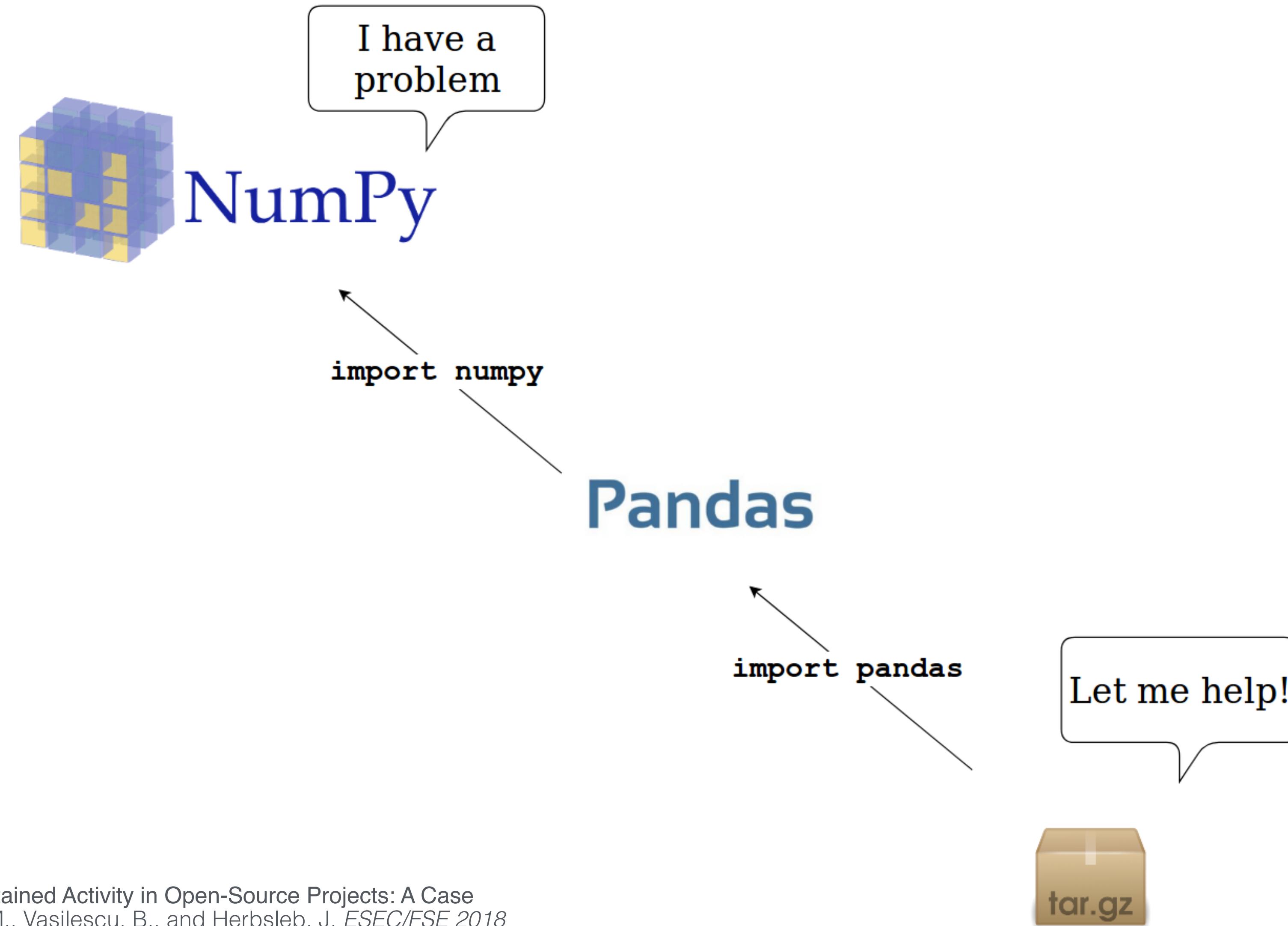
Interviews:

10 project maintainers



- Ecosystem-Level Determinants of Sustained Activity in Open-Source Projects: A Case Study of the PyPI Ecosystem. Valiev, M., Vasilescu, B., and Herbsleb, J. *ESEC/FSE 2018*

Transitive downstream dependencies are



- Ecosystem-Level Determinants of Sustained Activity in Open-Source Projects: A Case Study of the PyPI Ecosystem. Valiev, M., Vasilescu, B., and Herbsleb, J. *ESEC/FSE 2018*

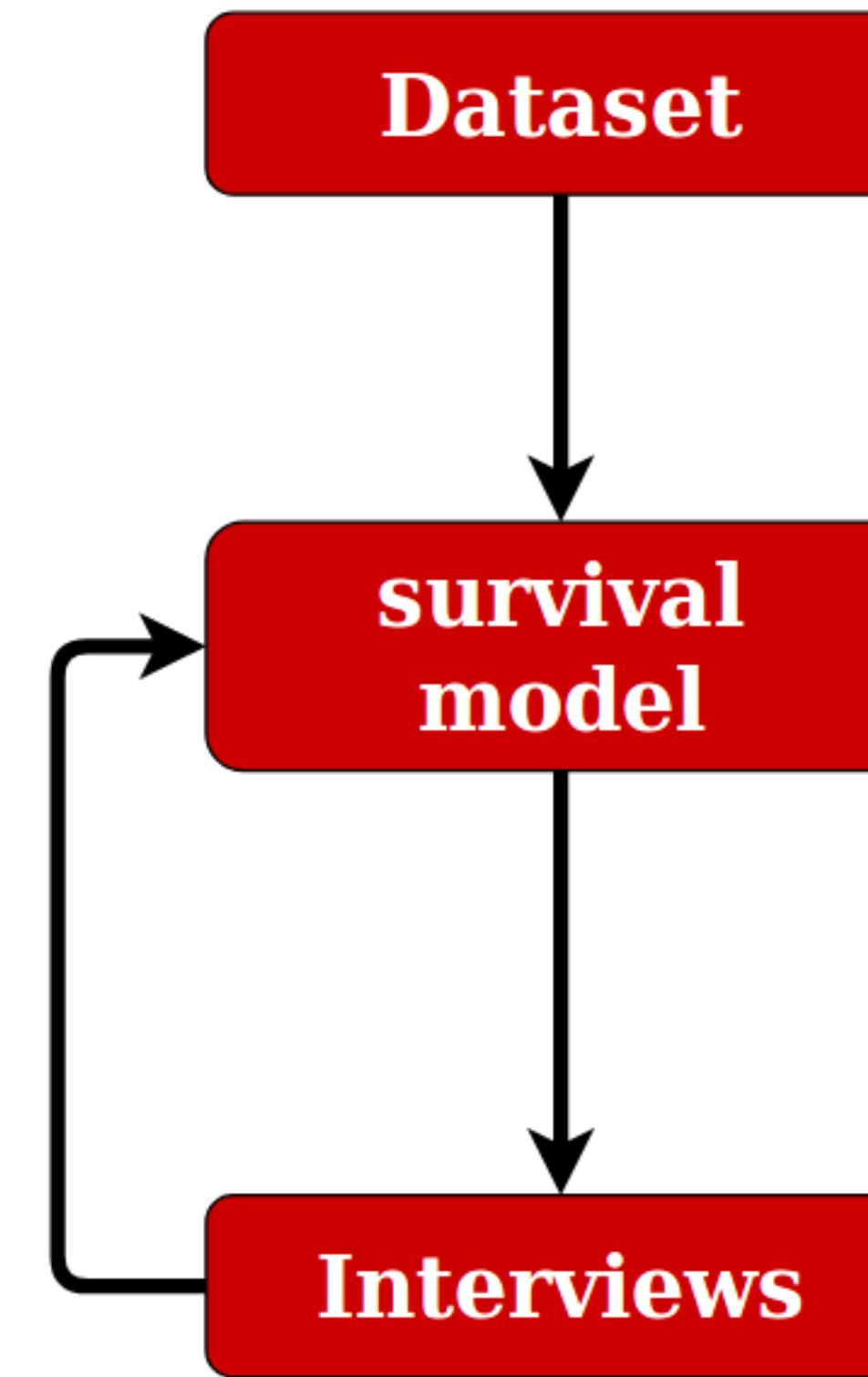
Transitive downstream dependencies are harmful

Feature: Katz centrality
(discounted transitive dependencies)

Early stage: **-12%** survival
Long term: **-27%**

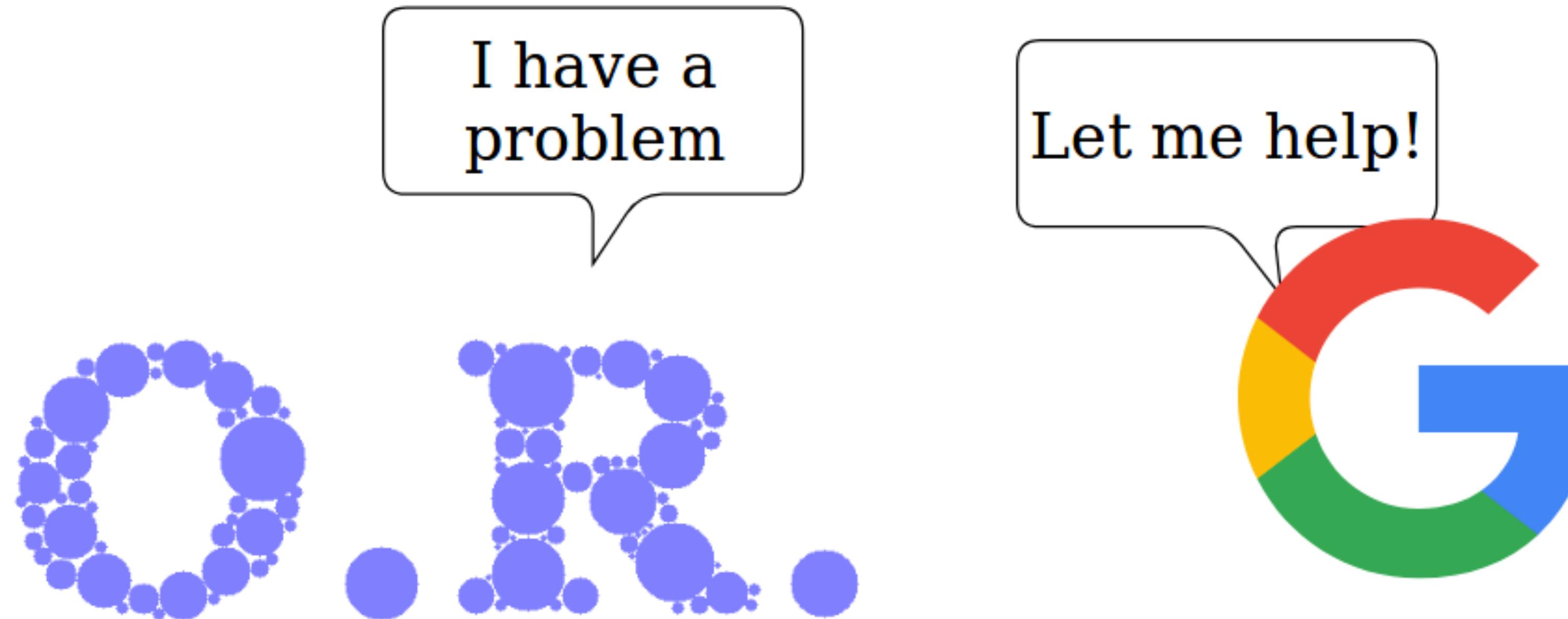
Interviews:

- less likely to fix
- just as likely to complain



- Ecosystem-Level Determinants of Sustained Activity in Open-Source Projects: A Case Study of the PyPI Ecosystem. Valiev, M., Vasilescu, B., and Herbsleb, J. *ESEC/FSE 2018*

Commercial involvement is



- Ecosystem-Level Determinants of Sustained Activity in Open-Source Projects: A Case Study of the PyPI Ecosystem. Valiev, M., Vasilescu, B., and Herbsleb, J. *ESEC/FSE 2018*

Commercial involvement is harmful

Feature:

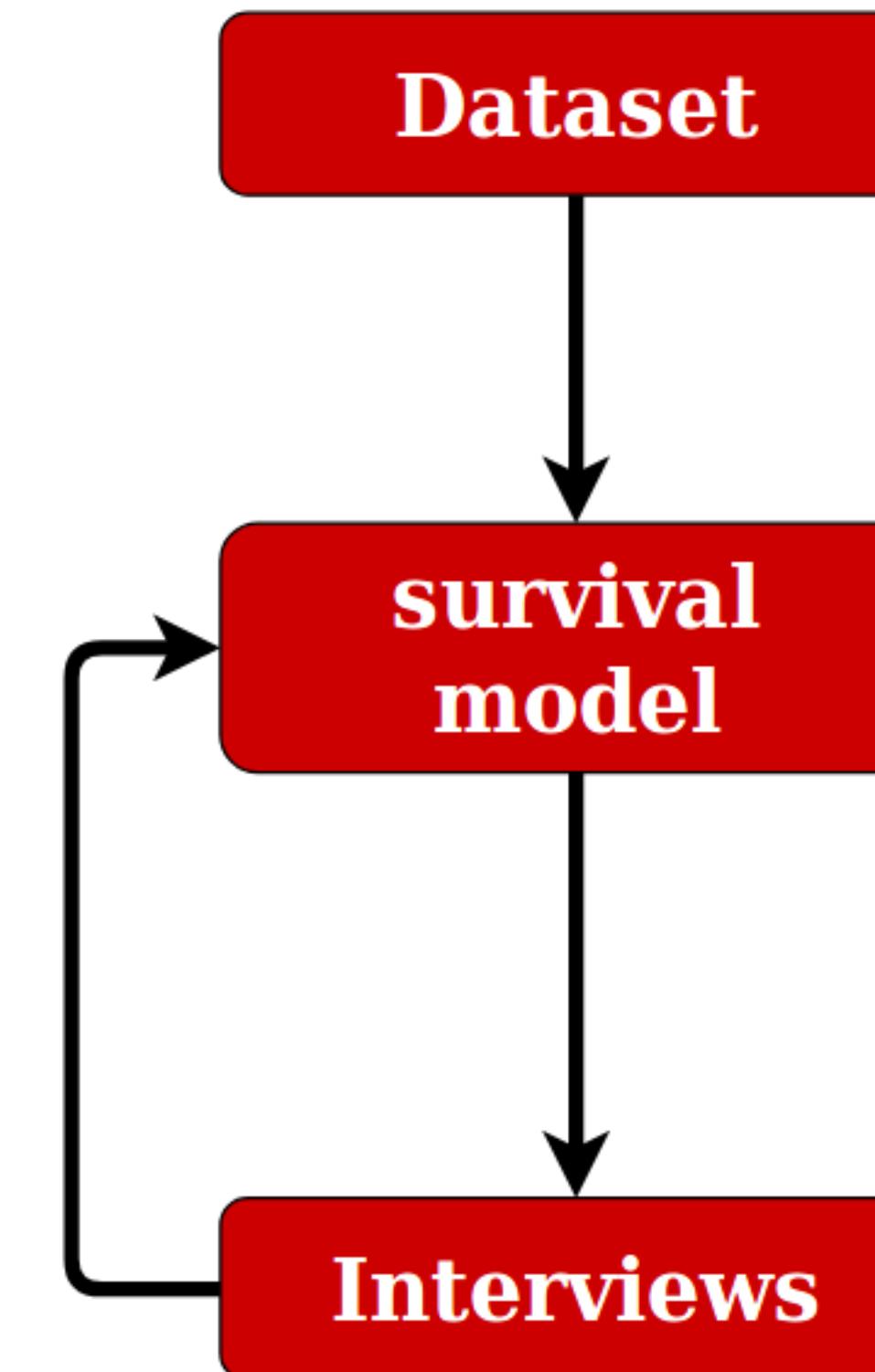
high commercial involvement

Early stage: **-51%** survival

Long term: **-15%**

Interviews:

- companies bring more resources
- but they can withdraw anytime



• Ecosystem-Level Determinants of Sustained Activity in Open-Source Projects: A Case Study of the PyPI Ecosystem. Valiev, M., Vasilescu, B., and Herbsleb, J. *ESEC/FSE 2018*

Example #1 conclusion: Ecosystem-level factors play an important role

New **signals** to display these otherwise unobservable ecosystem-level qualities:

- position in the network
- level of organizational support

The River of CPAN

Mon 20 April 2015

This blog post describes dependencies and reverse dependencies in CPAN. The river is Perl itself with all distributions in it. Every distribution contains all distributions that depend on it. Distributions sit somewhere in the river. Some distributions have many upstream dependencies, some have few. Some have reverse dependencies. There's a lot more.

Why a river?

If you pollute a river you might cause problems for everyone downstream of you. And you're relying on the distributions upstream of you not polluting the river.

For CPAN, the pollution is bugs: if one of your upstream dists has a buggy version released to CPAN, it might break your distribution, but it might not.

The further upstream a distribution, the more distributions it can potentially break, should it pollute the river.

So what?

CPAN authors / maintainers should know where their distributions sit on the river. We should help with that, and with visualising the upstream and downstream distributions. We should let authors know when a distribution moves up or down the river, particularly sudden large moves (if a distribution moves upstream and then a distribution much further upstream starts using your distribution, you zoom to a position upstream of them).

..... [MetaCPAN-Pod-XHTML-0.001002](#)
..... [Module-Reader-0.003003](#)
..... [Moo-2.003004](#)
..... [MooX-Aliases-0.001006](#)
..... [MooX-InsideOut-0.001004](#)
..... [MooX-Aliases-0.001006](#)

Example #2: “It’s most important that the people seem nice”

How do people choose which project to contribute to?

Interviews:

15 GitHub users

Data:

~10K npm packages

Model:

Logistic regression
(has new contributors)

The **tone of the community** is an important factor in both interviews and model.

maintainers polite

Asking for help explicitly is an important factor in the interviews.

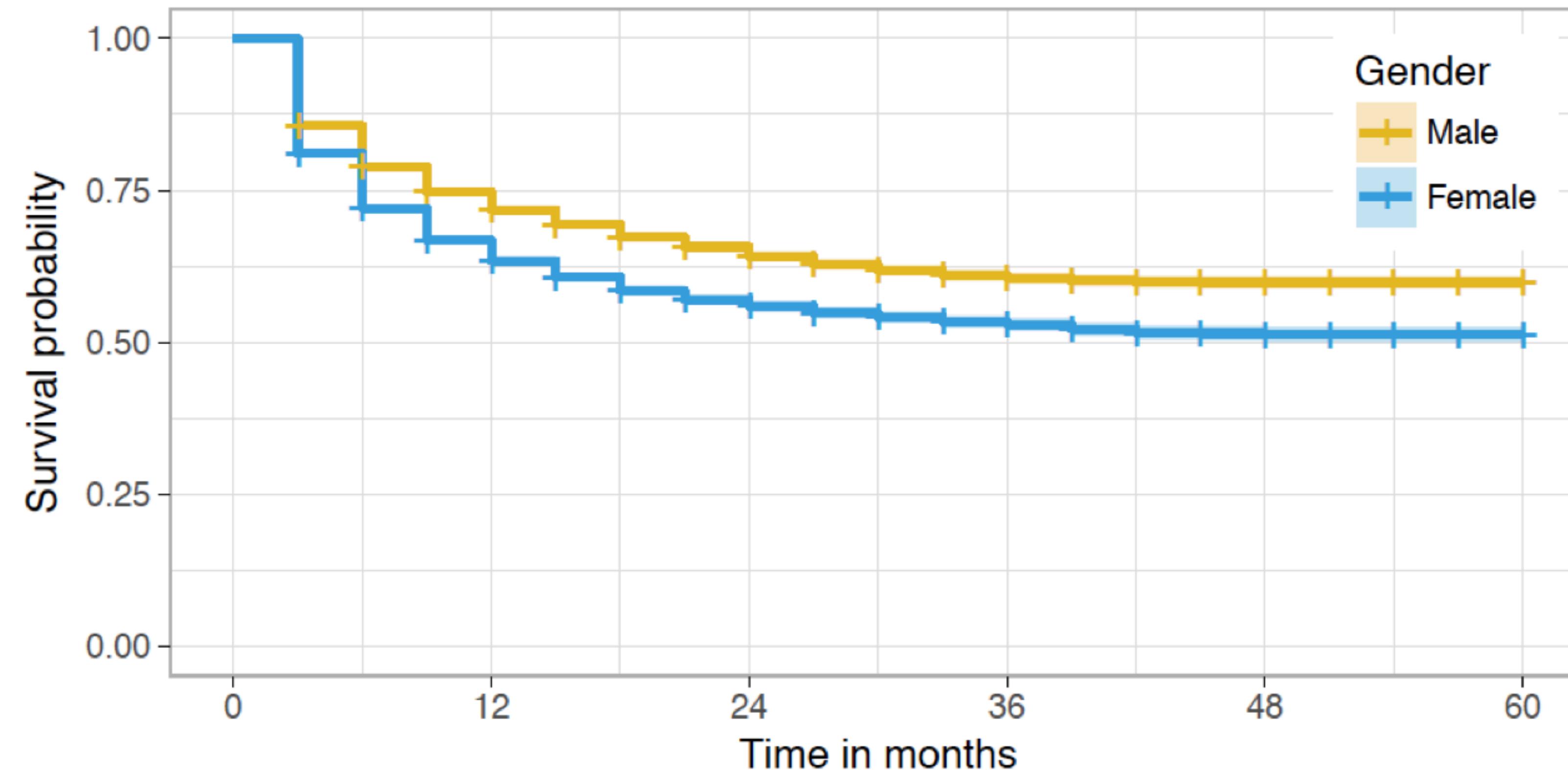
PRs welcome

help wanted

- The Signals that Potential Contributors Look for When Choosing Open-source Projects.
Qiu, S., Li, Yucen., Padala, S., Sarma, A., and Vasilescu, B. *Under review 2019*

Example #3: Building social capital

Why do women disengage earlier than men?



- Going Farther Together: The Impact of Social Capital on Sustained Participation in Open Source.
Qiu, H.S., Nolte, A., Brown, A., Serebrenik, A., and Vasilescu, B. *ICSE 2019*

Example #3: Building social capital

Why do women disengage earlier than men?

- Part of the explanation comes from the developer survey in our paper
- Reasons why people disengage:
 - Work-related (e.g., new job)
 - Personal* (e.g., different hobby)

**women cite more often than men*

- Going Farther Together: The Impact of Social Capital on Sustained Participation in Open Source.
Qiu, H.S., Nolte, A., Brown, A., Serebrenik, A., and Vasilescu, B. ICSE 2019

Why do People Give Up FLOSSing? A Study of Contributor Disengagement in Open Source

Courtney Miller^{1*}, David Widder², Christian Kästner², and Bogdan Vasilescu²
¹ New College of Florida, USA

Abstract. [REDACTED] open source that it is cri has also rev participate in ited research gage, and fac disengageme study, combi and predictiv find that dif for different some kind o also find that works on, wh and how muc disengageme

Why do developers take breaks from contributing to OSS projects? A preliminary analysis

Giuseppe Iaffaldano
Department of Computer Science
University of Bari
Bari, Italy
giuseppe.iaffaldano@uniba.it

Marco Gerosa
School of Informatics, Computing, and Cyber Systems
Northern Arizona University
marco.gerosa@nau.edu

Igor Steinmacher
School of Informatics, Computing, and Cyber Systems
Northern Arizona University
igor.steinmacher@nau.edu

Filippo Lanubile
Department of Computer Science
University of Bari
Bari, Italy
filippo.lanubile@uniba.it

See also:

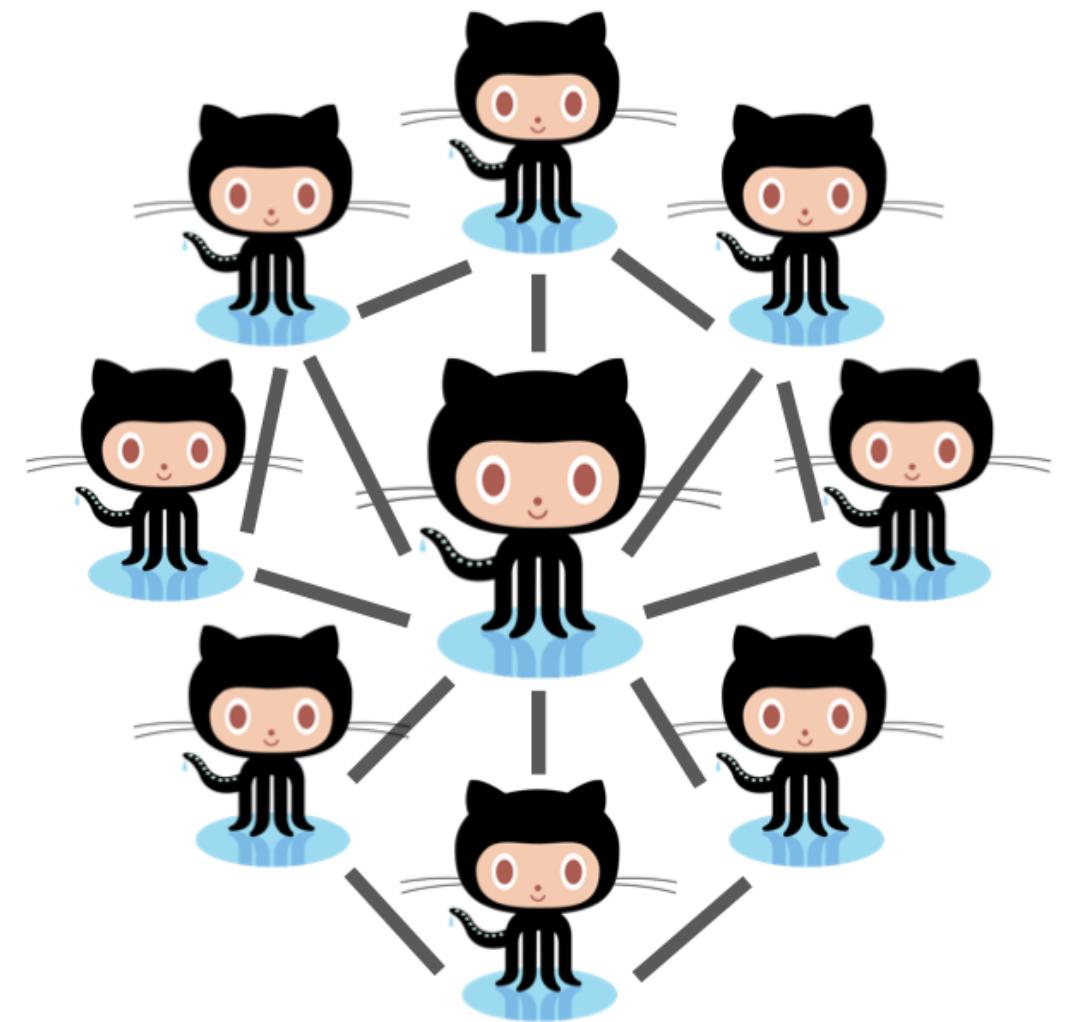
community [19]. On the other hand, it may disrupt the community and lower the product quality [15], [20].

By lurking in some projects on GitHub, we noticed that some developers take long breaks from development, while others suddenly disappear from the contribution timeline. We came up with metaphors suggesting that developers may spend some time sleeping or they can die. So, in this position paper, we explore the phenomenon of developers becoming inactive or abandoning the projects. To do so, we introduce the concepts of *sleeping* and *dead* developers, representing those developers who take temporary or permanent breaks from contributing code to the projects.

With this position paper, we want to open a discussion around this topic and bring evidence of the reasons why developers leave the projects and of the signals to help to identify that this phenomenon is happening.

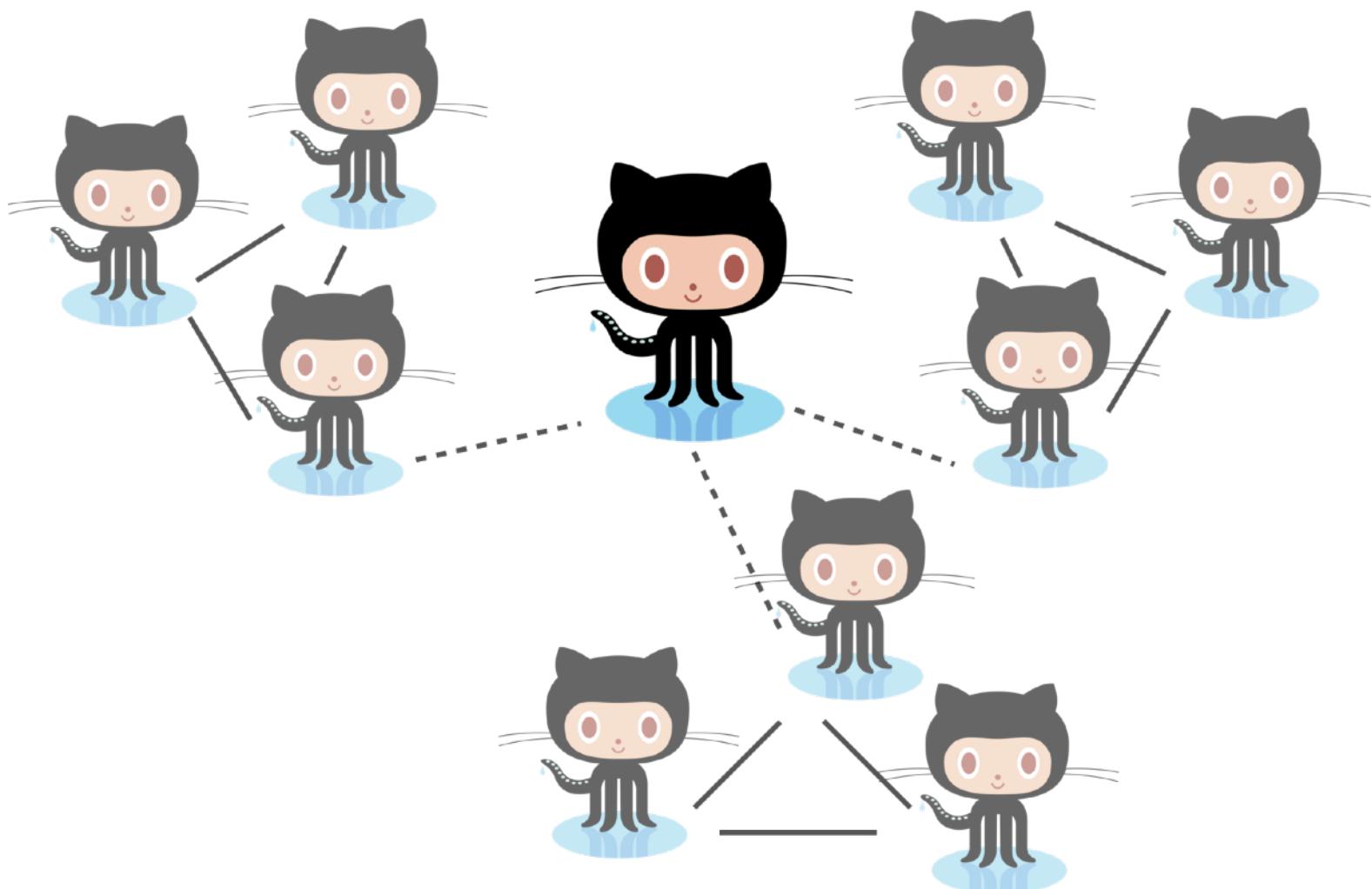
Social capital theory explains long-term engagement

Bonding social capital:
benefiting from strongly
connected network



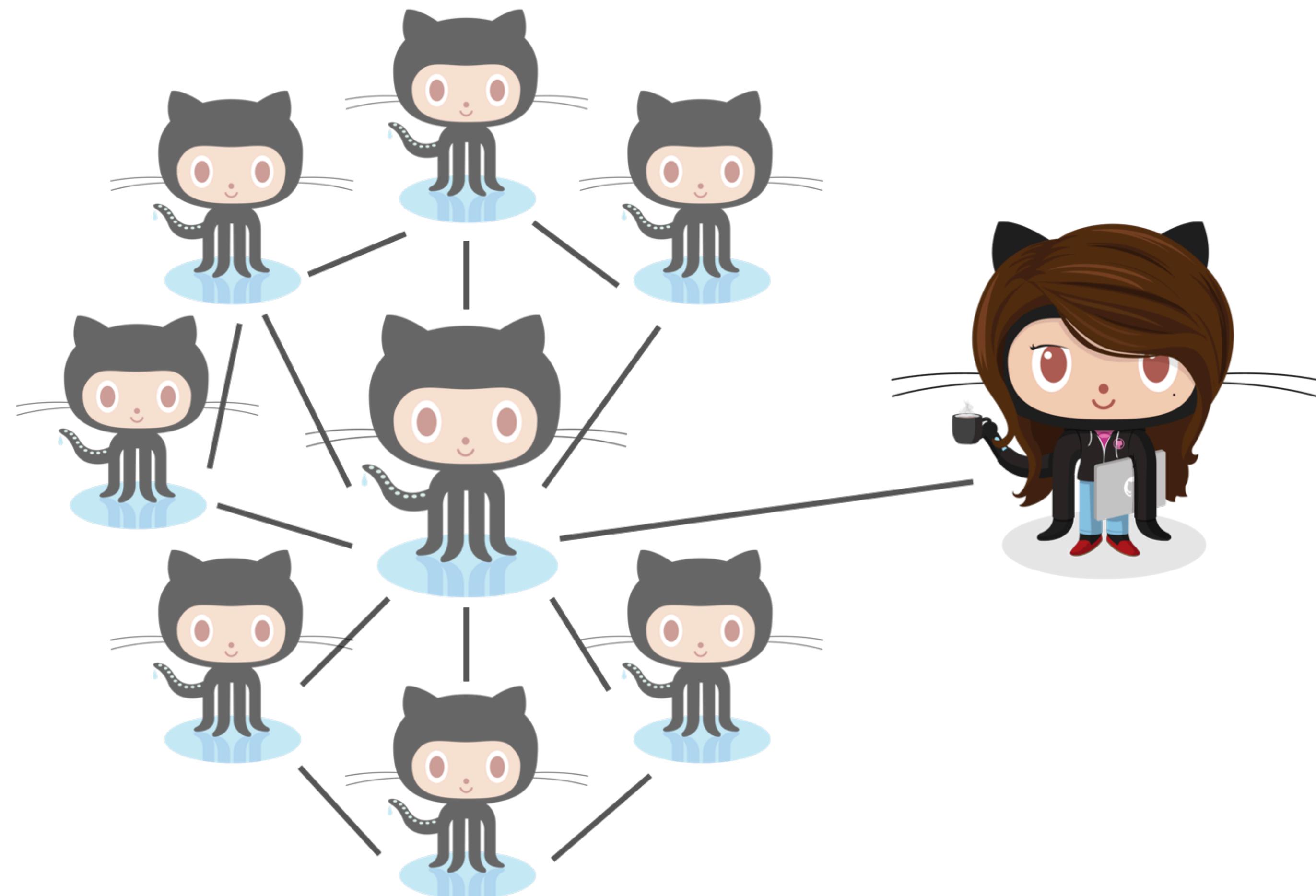
Willingness to continue
(Coleman, 1990)

Bridging social capital:
benefiting from network with
diverse info



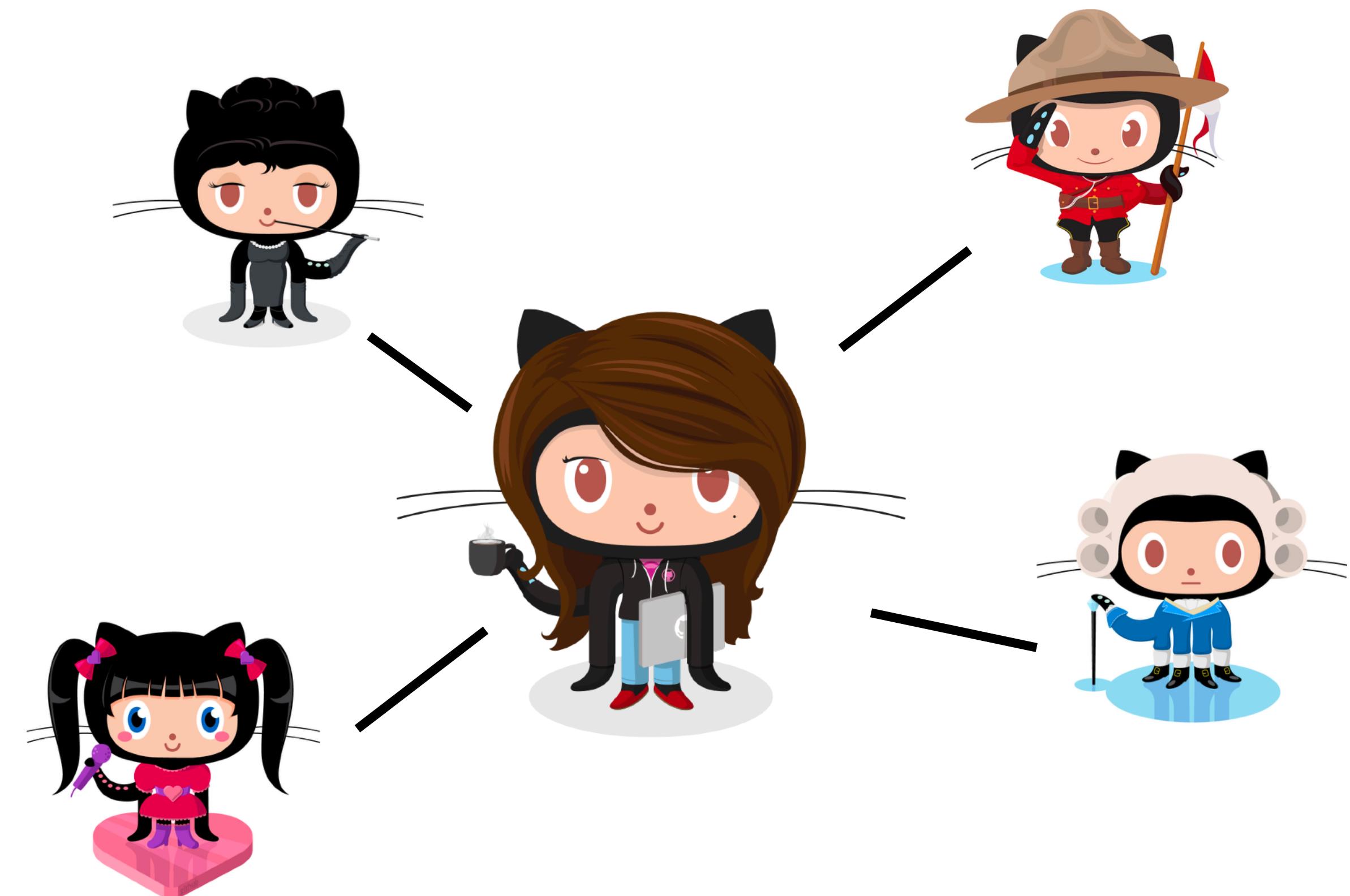
Opportunity to continue
(Burt, 1998, 2001)

Cohesive networks might foster discrimination / exclusion

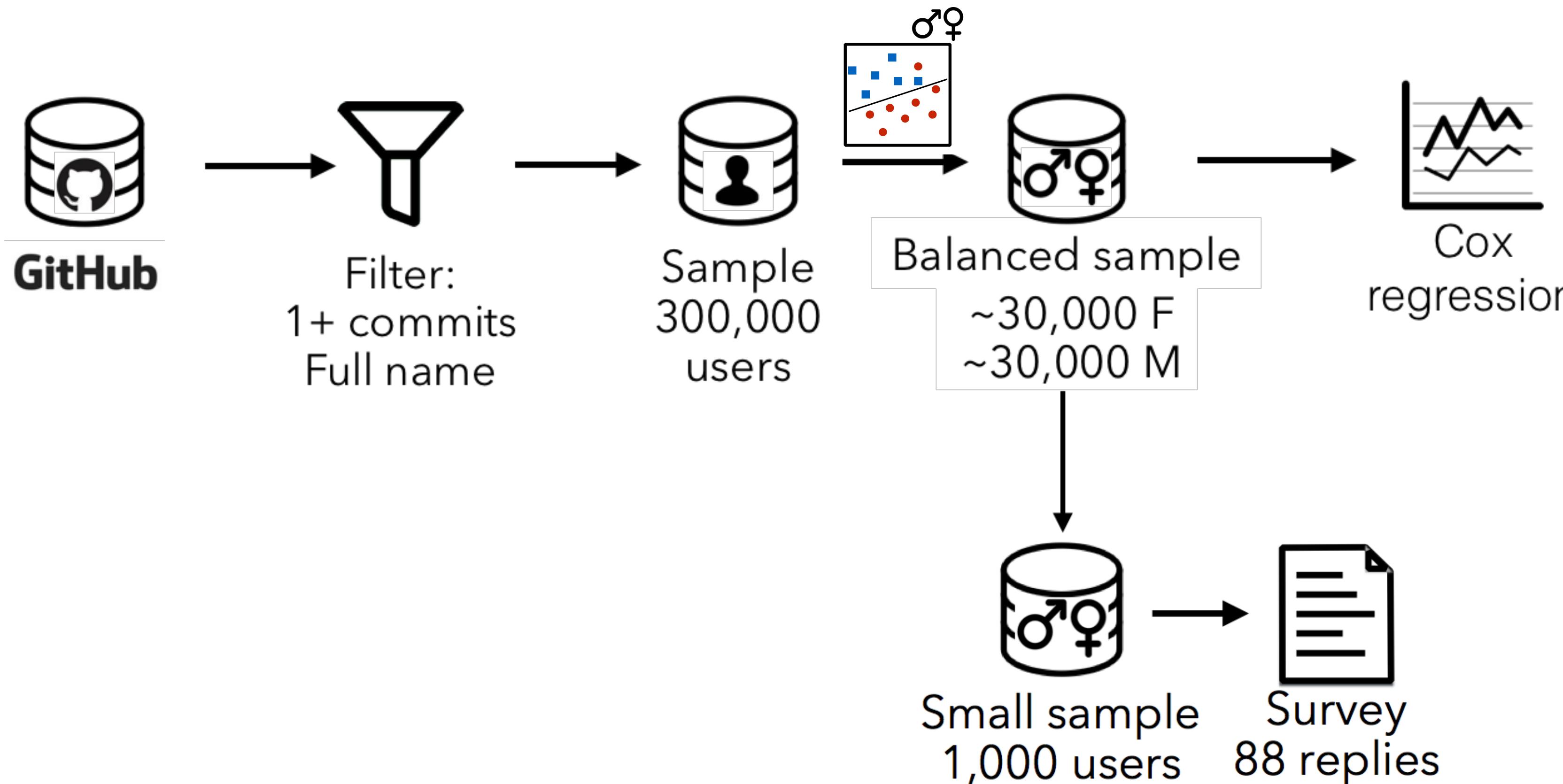


Being part of teams with more diverse information ~ more prolonged engagement, esp. for women

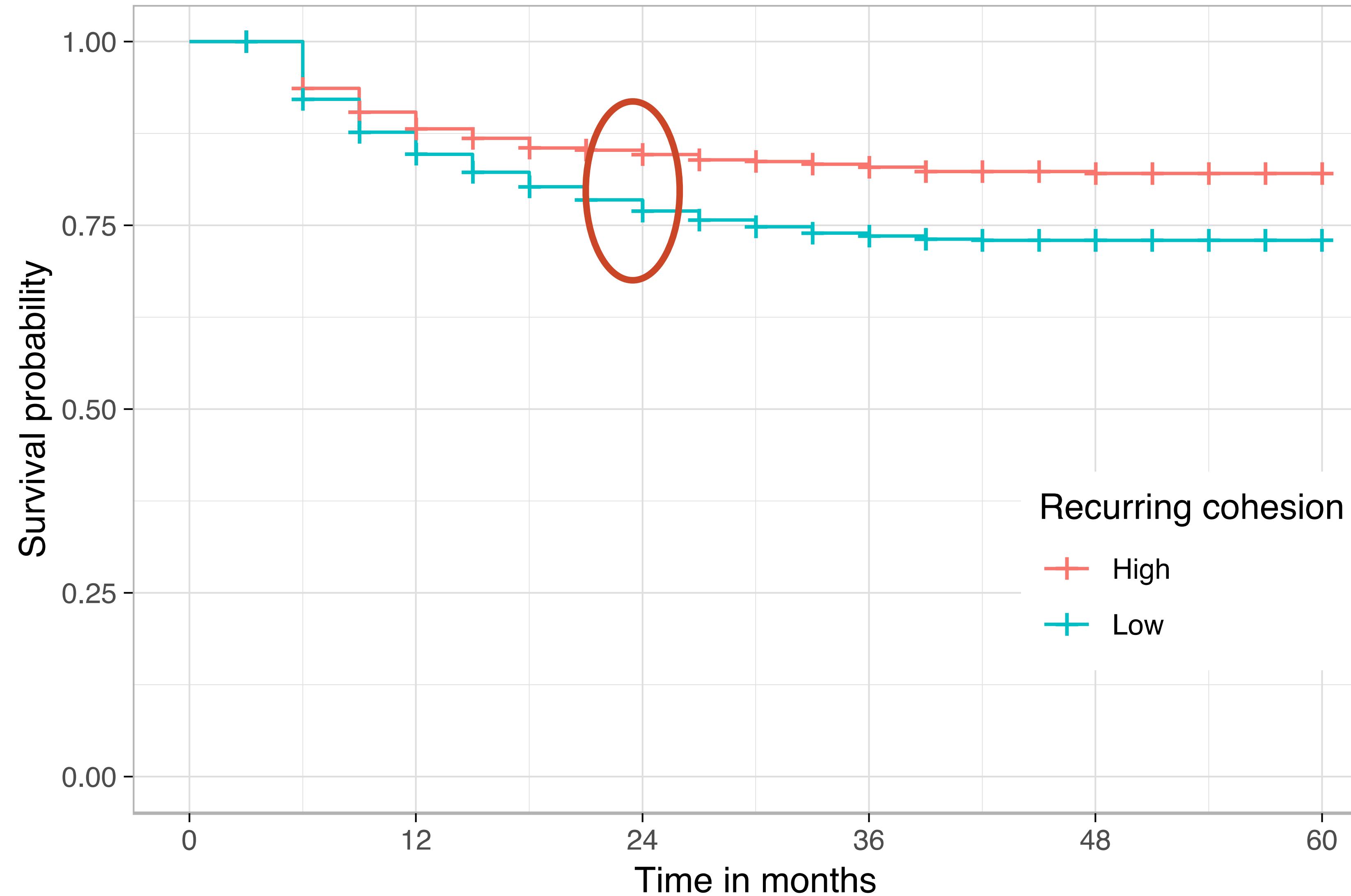
Information diversity should
reduce the risk of demographic-
based echo chambers.



Large-scale mixed-methods study

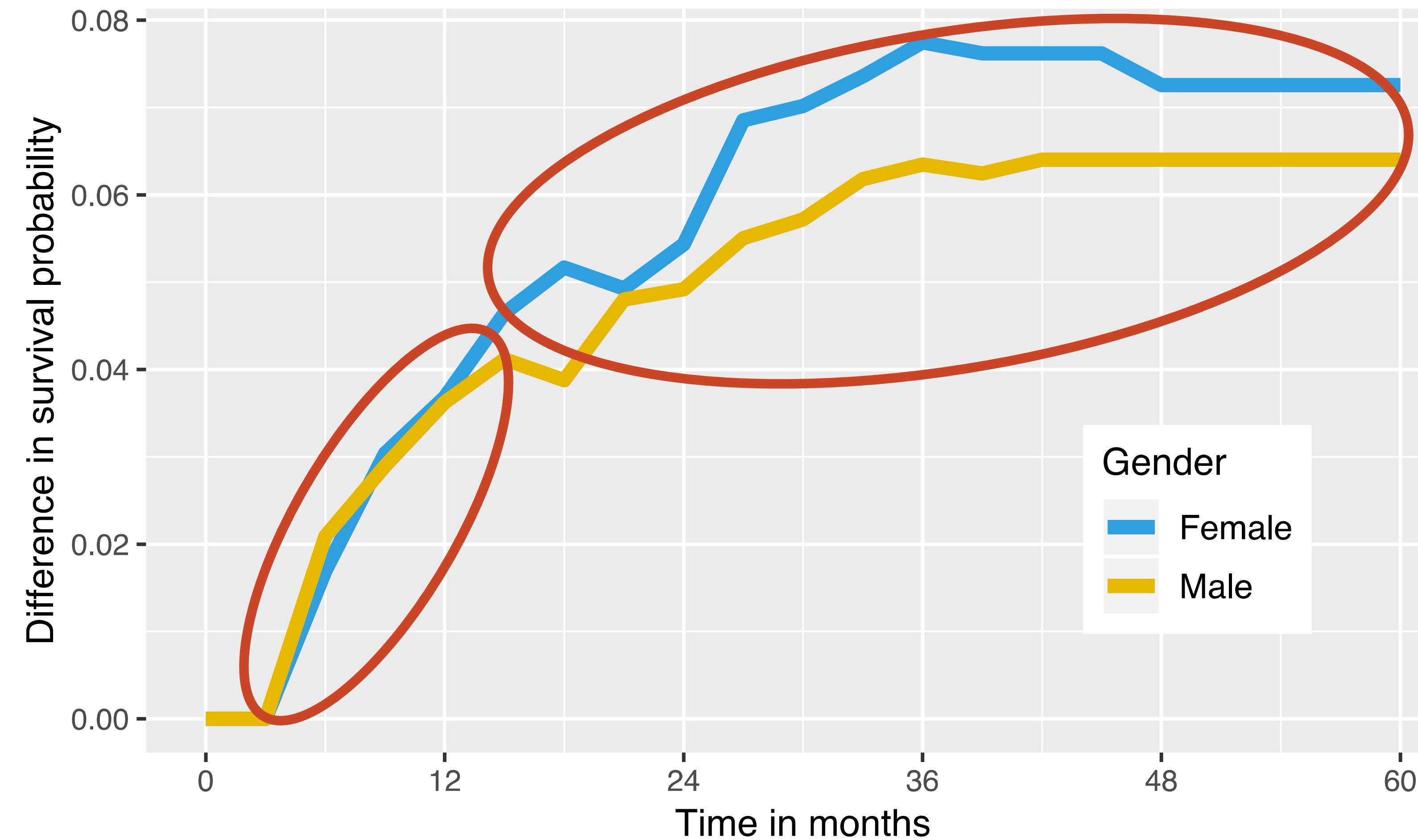


More social capital ~ more prolonged engagement

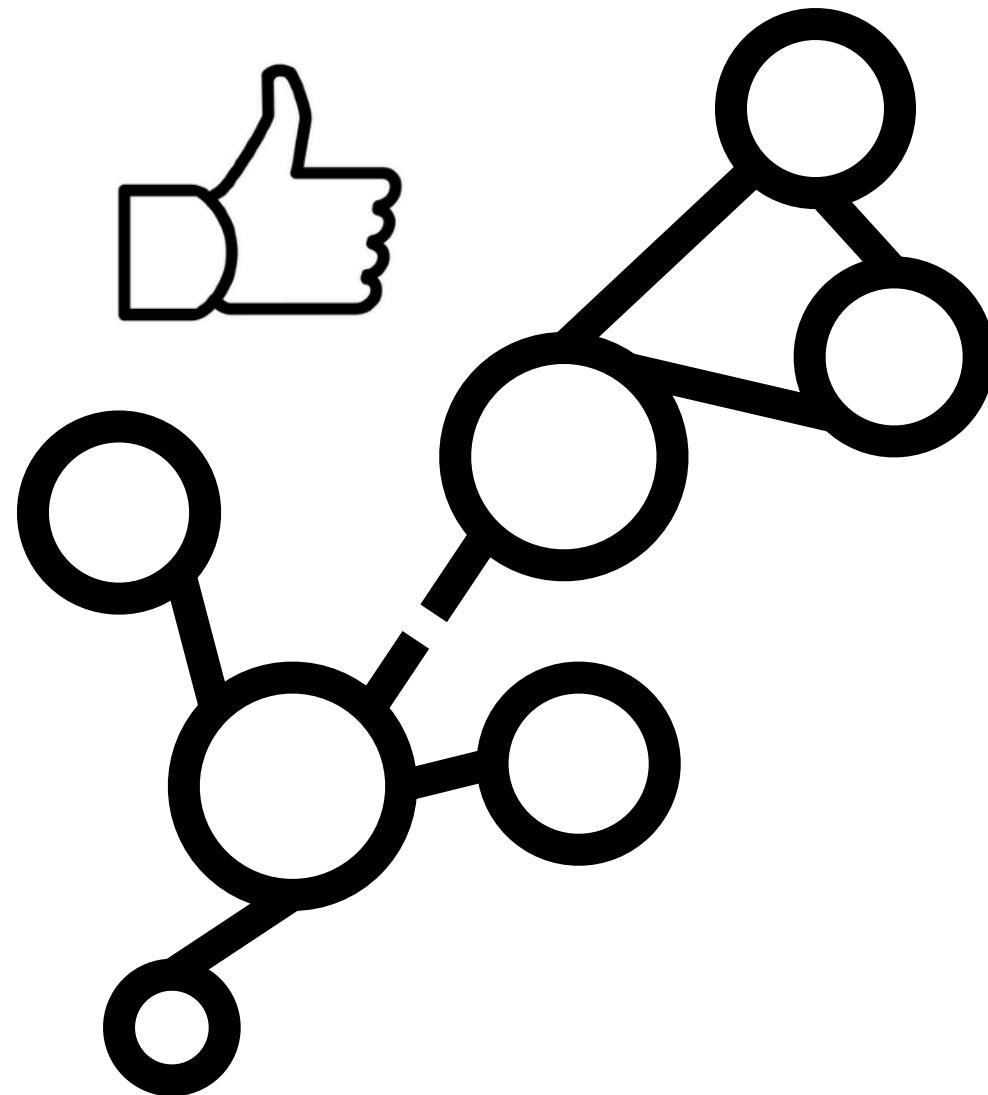


Women in language- (informationally-) diverse teams disengage at lower rates

Survival difference between contributors with high and low language diversity



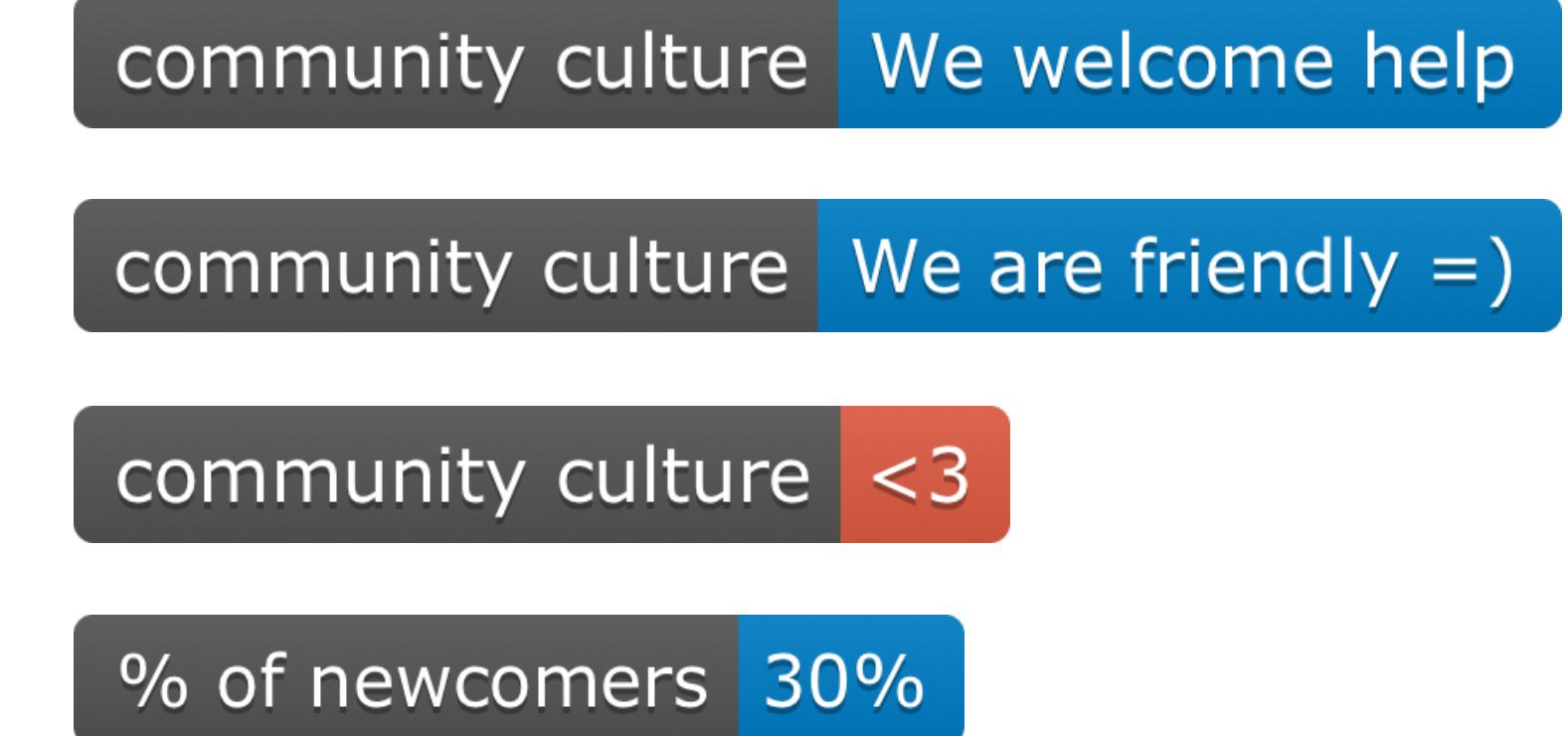
Example #3 conclusion:



Recommend projects that can help build social capital



Find relevant mentorship



Signal social capital moderators

Leveraging signals

Transparency is already a defining characteristic of the environment

This screenshot shows a GitHub profile page. At the top right, there are buttons for Contributions, Repositories, Public activity, Follow, and a dropdown menu. Below this, the profile picture is a cartoon cat holding a laptop with 'CV' on it. The profile name is 'npm, inc'. On the left, there's a sidebar with icons for npm, location, email, and GitHub. It also shows the user joined on Oct 31, 2011, with 776 Followers, 38 Starred, and 15 Following. Under 'Organizations', there are links to various organizations represented by icons.

Popular repositories

- breakfast-repo** (208 stars)
- x86-kernel** (48 stars)
- jsconf-2015-deck** (32 stars)
- ratpack** (32 stars)

Repositories contributed to

- npm/docs** (44 stars)
- mozilla/publish.webmaker.org** (2 stars)
- npm/marky-markdown** (104 stars)
- artisan-tattoo/assistant-frontend** (5 stars)
- npm/npm-camp** (1 star)

Public contributions

Summary of pull requests, issues opened, and commits. Learn how we count contributions.

Less More

Contributions in the last year: 1,886 total (Jan 24, 2015 – Jan 24, 2016)

Longest streak: 37 days (October 7 – November 12)

Current streak: 7 days (January 18 – January 24)

This screenshot shows the GitHub repository page for 'caolan / async'. At the top right, there are buttons for Watch (721), Star (23,937), Fork (2,203), and a dropdown menu. The repository name is 'caolan / async'. Below this, there are tabs for Code, Issues (21), Pull requests (6), Projects (0), Wiki, and Insights. The 'Code' tab is selected. The repository description is 'Async utilities for node and the browser <http://caolan.github.io/async/>'. It lists technologies: javascript, async, callbacks. Key metrics include 1,629 commits, 11 branches, 72 releases, and 206 contributors. The license is MIT. A large blue 'async' logo is prominently displayed. Below the logo, there are badges for build (passing), npm (v2.6.0), coverage (99%), gitter, join chat, examples (26348), jsDelivr, and hits/month (407k).

caolan / async

Watch 721 | Star 23,937 | Fork 2,203

Code Issues 21 Pull requests 6 Projects 0 Wiki Insights

Async utilities for node and the browser <http://caolan.github.io/async/>

javascript async callbacks

1,629 commits 11 branches 72 releases 206 contributors MIT

README.md

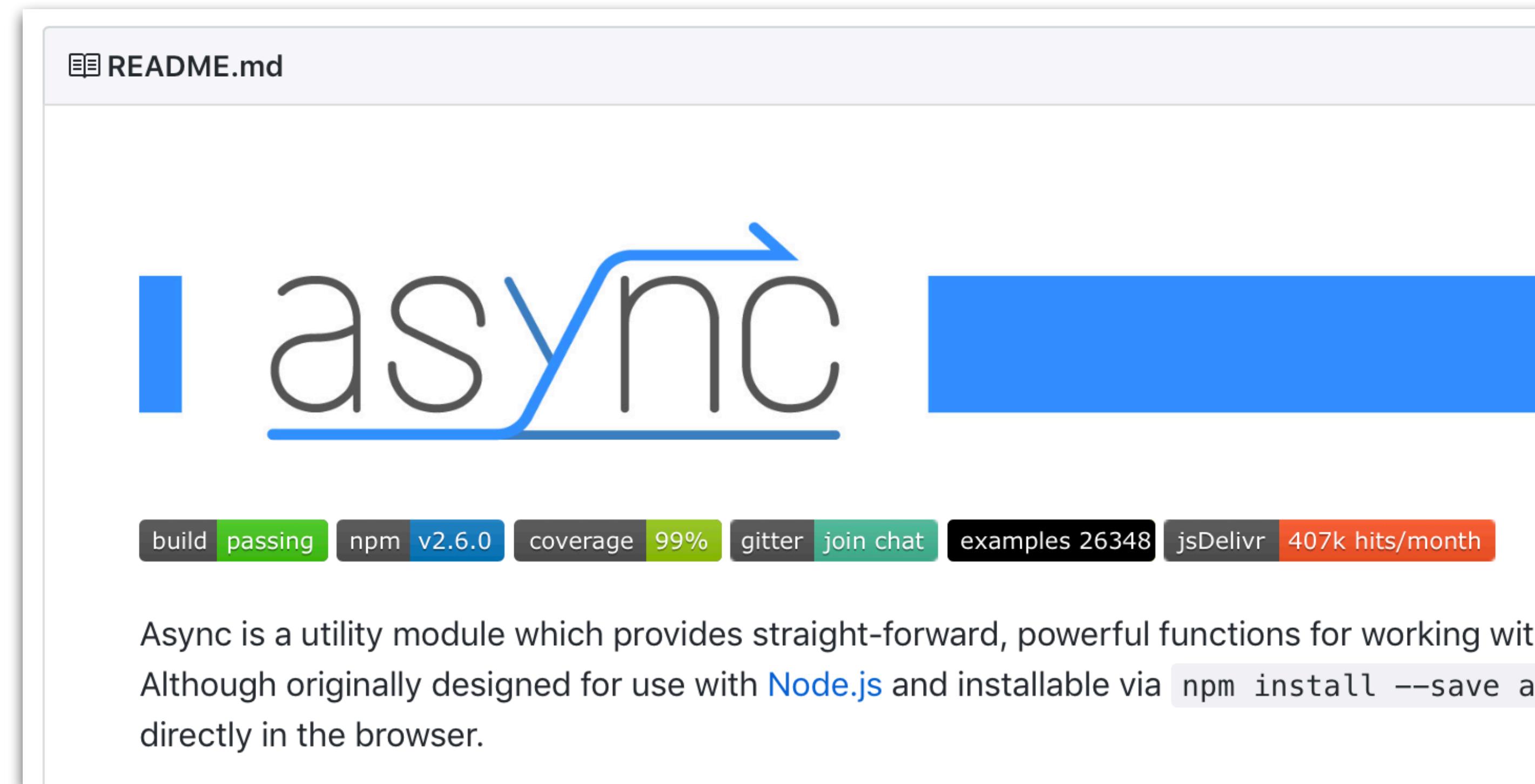
async

build passing npm v2.6.0 coverage 99% gitter join chat examples 26348 jsDelivr 407k hits/month

Async is a utility module which provides straight-forward, powerful functions for working with [asynchronous JavaScript](#). Although originally designed for use with [Node.js](#) and installable via `npm install --save async`, it can also be used directly in the browser.

Signals are customizable

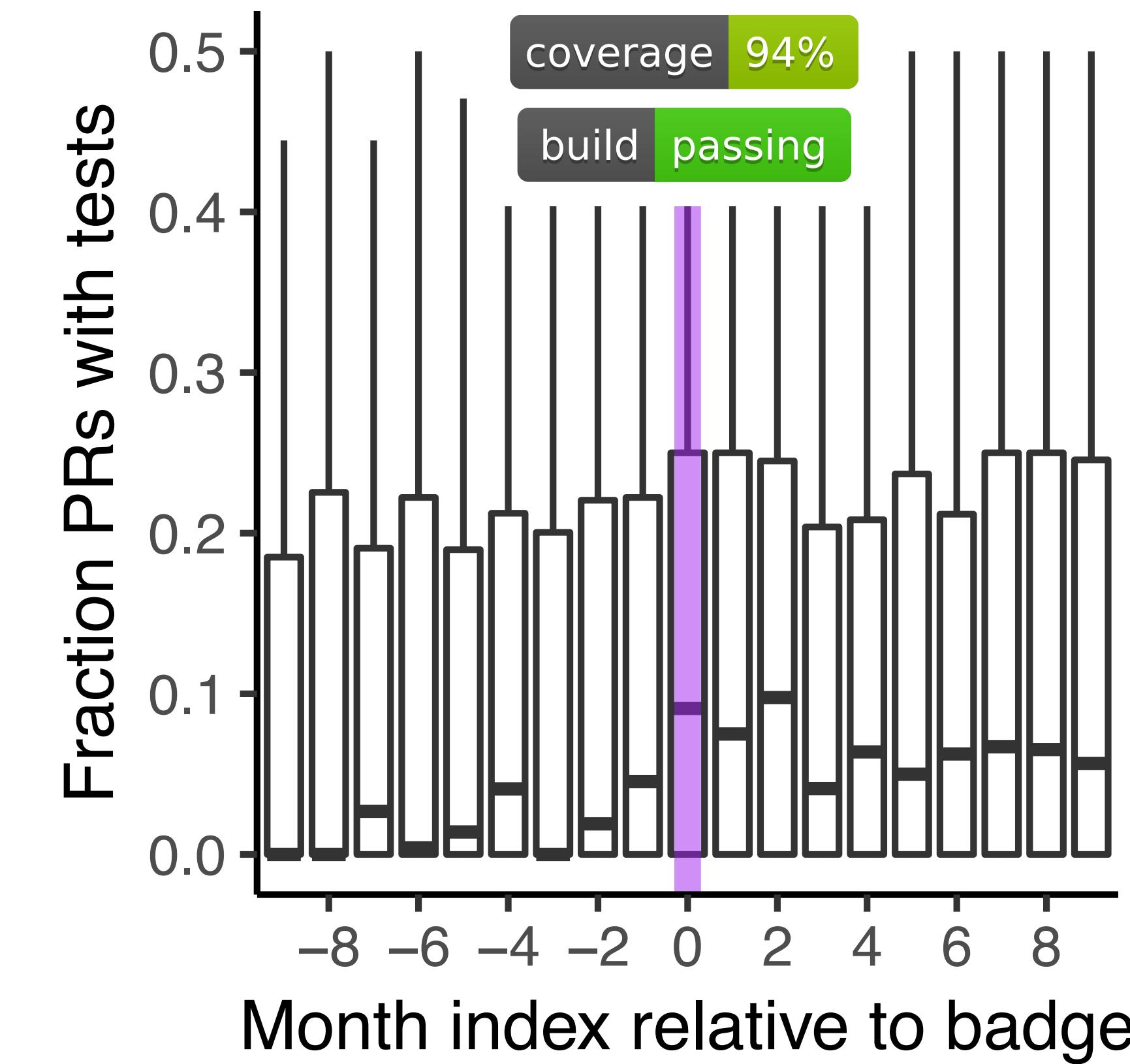
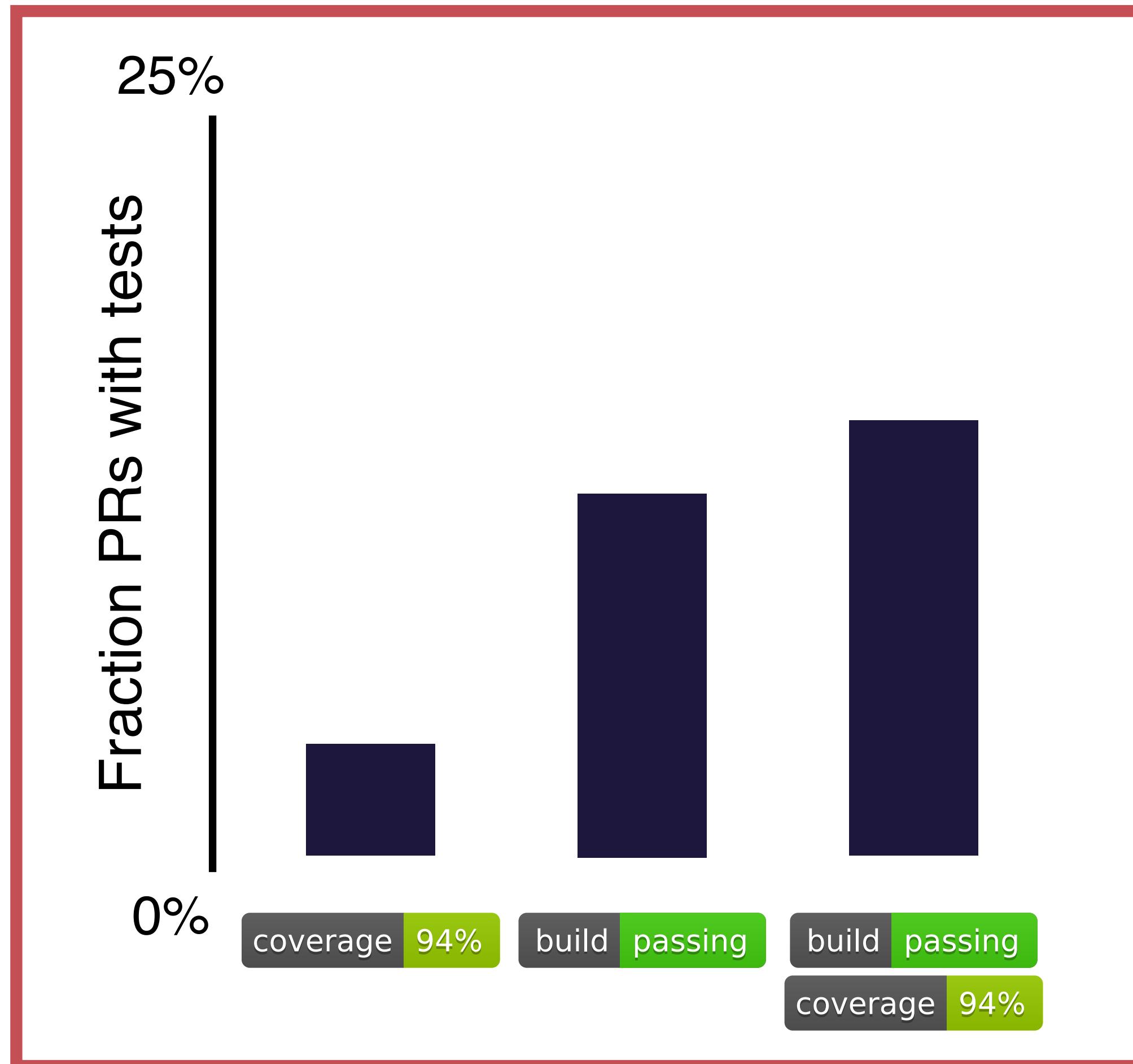
- E.g., repository badges



- Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the npm Ecosystem. Trockman, A., Zhou, S., Kästner, C., and Vasilescu, B. *ICSE 2018*

Signals are effective at steering behavior

build passing + coverage 94% badges indicate more tests in PRs



- Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the npm Ecosystem. Trockman, A., Zhou, S., Kästner, C., and Vasilescu, B. ICSE 2018

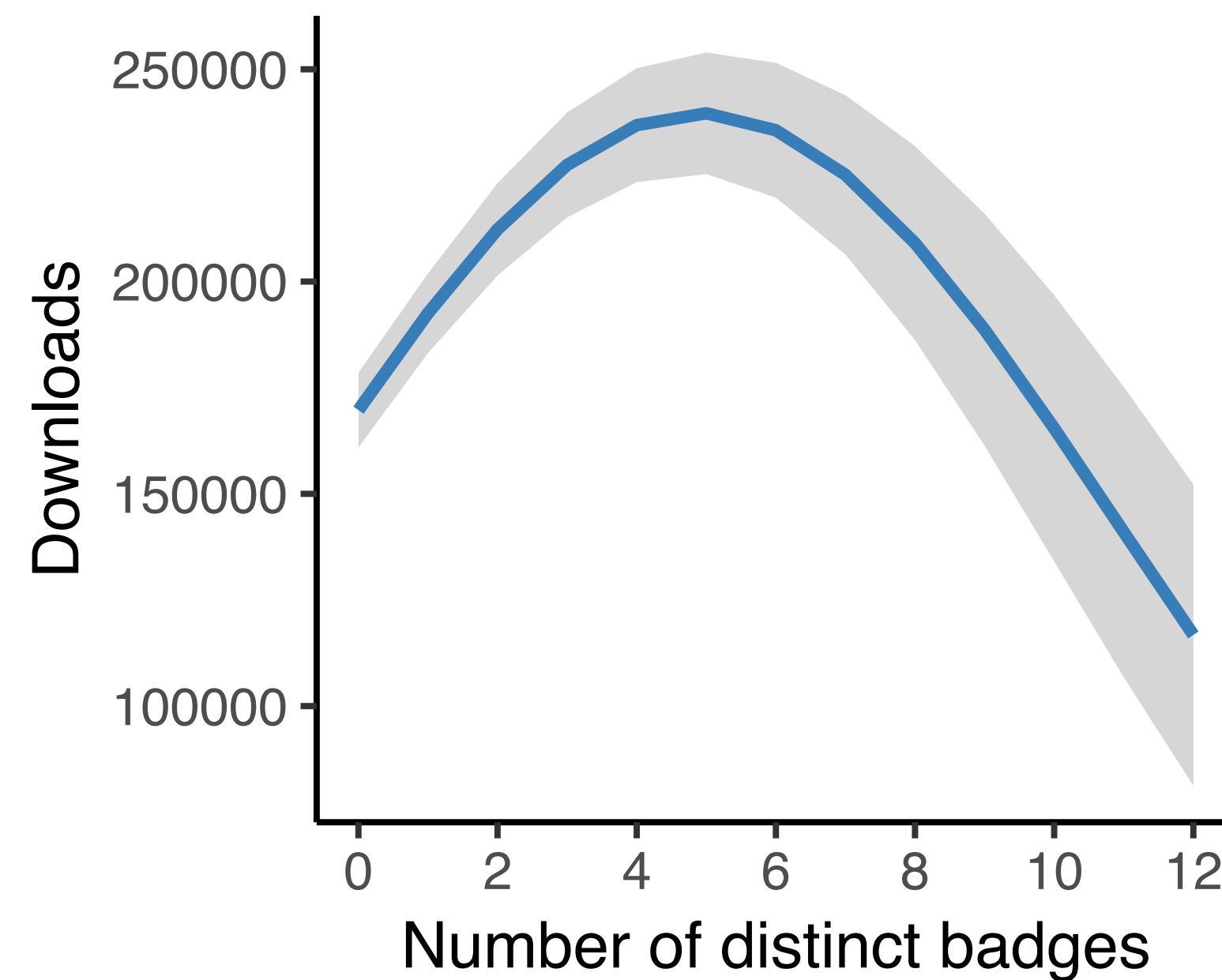
Hypothesis: Signals can help the ecosystem self-regulate





Signals are no panacea

Attractiveness wears off beyond 5 badges



Developers are aware of each other's gender

Which of the following characteristics of your team members are you aware of?

- 74% • Programming skills
- 48% • **Gender**
- 45% • Real name
- 42% • Social skills
- 40% • Country of residence
- 39% • Personality
- 31% • Reputation as programmer
- 30% • Ethnicity
- 30% • Employment
- 28% • GitHub experience
- 26% • Educational level
- 23% • Age
- 11% • Hobbies
- 4% • Political views

• Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the npm Ecosystem. Trockman, A., Zhou, S., Kästner, C., and Vasilescu, B. *ICSE 2018*

• Perceptions of Diversity on GitHub: A User Survey. Vasilescu, B., Filkov, V., and Serebrenik, A. *CHASE 2015*

“Sexist behavior in F/LOSS is as constant as it is extreme”

Article



‘Patches don’t have gender’: What is not open in open source software

new media & society
14(4) 669–683
© The Author(s) 2011
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1461444811422887
nms.sagepub.com
SAGE

Dawn Nafus
Intel Labs, USA

Abstract

While open source software development promises a fairer, more democratic model of software production often compared to a gift economy, it also is far more male dominated than other forms of software production. The specific ways F/LOSS instantiates notions of openness in everyday practice exacerbates the exclusion of women. ‘Openness’ is a complex construct that affects more than intellectual property arrangements. It weaves together ideas about authorship, agency, and the circumstances under which knowledge and code can and cannot be exchanged. While open source developers believe technology is orthogonal to the social, notions of openness tie the social to the technical by separating persons from one another and relieving them of obligations that might be created in the course of other forms of gift exchange. In doing so, men monopolize code authorship and simultaneously de-legitimize the kinds of social ties necessary to build mechanisms for women’s inclusion.

“I have used a fake GitHub handle [...] so that people would assume I was male”

Article



new media & society

‘Patches don’t have gender. What is not open in open source software’

Dawn Nafus
Intel Labs, USA

Abstract

While open source software development promises a more democratic form of software production often compared to a gift economy, the reality is more complex than other forms of software production. The specific ways in which the ideal of openness in everyday practice exacerbates the exclusions of the open source construct that affects more than intellectual property rights, such as ideas about authorship, agency, and the circumstances under which ideas can and cannot be exchanged. While open source development is often seen as tied to the social, notions of openness tie the social to the technical, and one another and relieving them of obligations that may be associated with other forms of gift exchange. In doing so, men monopolize the kinds of social ties necessary to build and maintain open source projects.

Perceptions of Diversity on GitHub: A User Survey

Bogdan Vasilescu
University of California, Davis
vasilescu@ucdavis.edu

Vladimir Filkov
University of California, Davis
filkov@cs.ucdavis.edu

Alexander Serebrenik
Eindhoven University of Technology
a.serebrenik@tue.nl

Abstract—Understanding one’s work environment is important for one’s success, especially when working in teams. In virtual collaborative environments this amounts to being aware of the technical and social attributes of one’s team members. Focusing on Open Source Software teams, naturally very diverse both socially and technically, we report the results of a user survey that tries to resolve how teamwork and individual attributes are perceived by developers collaborating on GITHUB, and how those perceptions influence their work. Our findings can be used as complementary data to quantitative studies of developers’ behavior on GITHUB.

I. INTRODUCTION

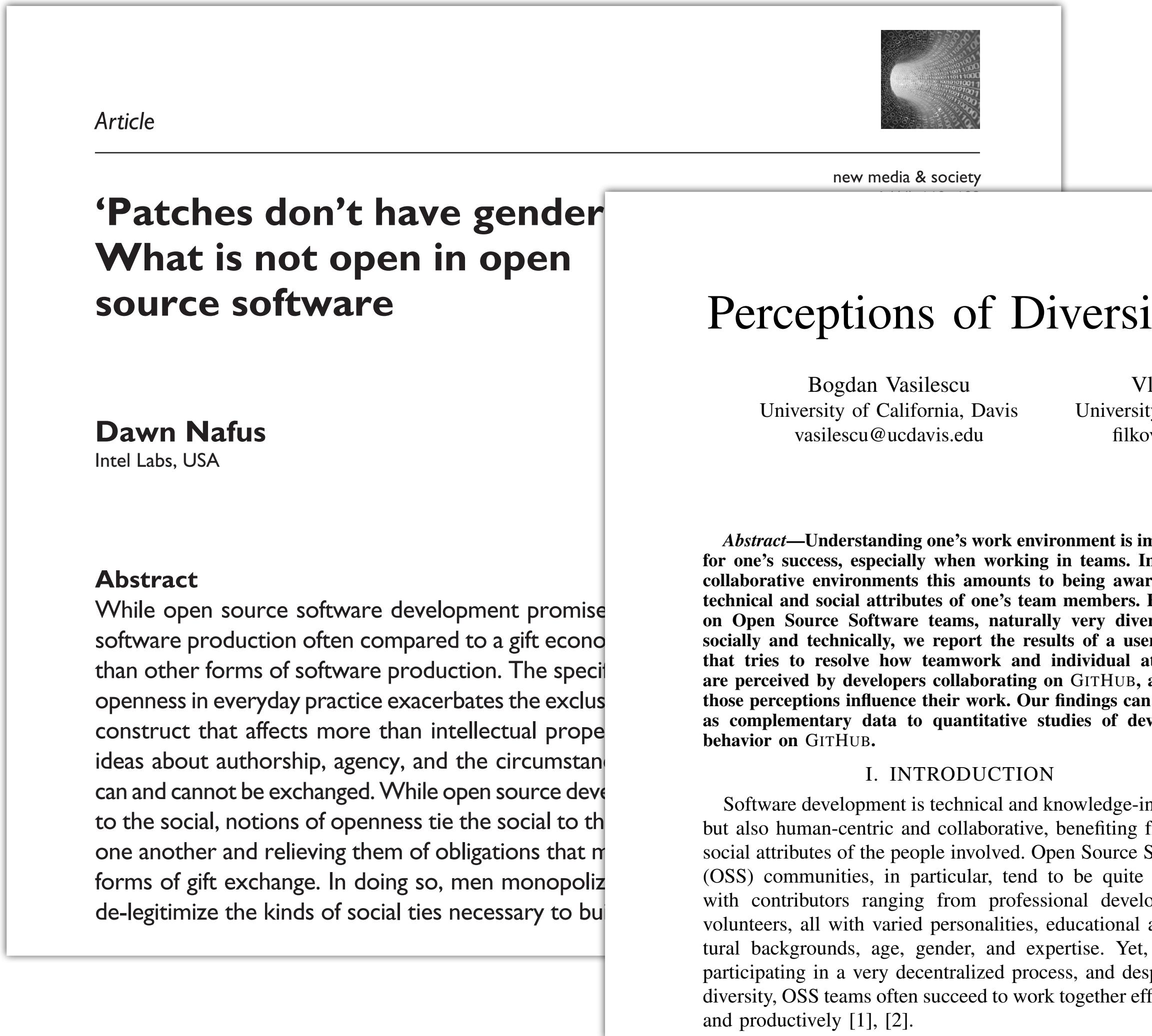
Software development is technical and knowledge-intensive, but also human-centric and collaborative, benefiting from the social attributes of the people involved. Open Source Software (OSS) communities, in particular, tend to be quite diverse, with contributors ranging from professional developers to volunteers, all with varied personalities, educational and cultural backgrounds, age, gender, and expertise. Yet, despite participating in a very decentralized process, and despite this diversity, OSS teams often succeed to work together effectively and productively [1], [2].

attributes (e.g., gender, tenure, political views) on the overall work environment. Our previous study [7] was, to the best of our knowledge, the first to consider effects of gender diversity on productivity and turnover in OSS communities, and one of the very few studies of diversity in general in OSS or other online peer production systems (e.g., [14]–[16]).

In this paper we offer a qualitative perspective of diversity in software teams: we report the results of a user survey that tries to resolve how teamwork and individual attributes are perceived by developers collaborating on GITHUB, and how those perceptions influence their work. We address a number of research questions, as discussed next.

OSS teams are typically more fluid and less tangible than their offline counterparts. They tend to form and dissolve organically around the task at hand, facing high turnover [17], while interactions between members are often limited to online channels [18]. In addition, GITHUB’s implementation of the pull-based development model [19] enables anyone to submit changes to any repository with minimal effort, through pull requests (the so-called “drive-by” commits [13]). We wish to understand whether this unprecedented low barrier to entry for

Pull request acceptance rates are lower when gender is apparent



Article

'Patches don't have gender. What is not open in open source software'

Dawn Nafus
Intel Labs, USA

Abstract
While open source software development promises a more collaborative and meritocratic way of producing software, it has been argued that the social dynamics of open source software production often compare to a gift economy rather than other forms of software production. The specific nature of openness in everyday practice exacerbates the exclusions that construct a social hierarchy that affects more than intellectual property rights. Ideas about authorship, agency, and the circumstances under which ideas can and cannot be exchanged. While open source development is often seen as being more open to the social, notions of openness tie the social to the technical and the individual to one another and relieving them of obligations that may be associated with other forms of gift exchange. In doing so, men monopolize the kinds of social ties necessary to build OSS teams.

Perceptions of Diversity

Bogdan Vasilescu
University of California, Davis
vasilescu@ucdavis.edu

Vlad Filkov
University of California, Davis

Abstract—Understanding one's work environment is important for one's success, especially when working in teams. In collaborative environments this amounts to being aware of the technical and social attributes of one's team members. Focusing on Open Source Software teams, naturally very diverse both socially and technically, we report the results of a user study that tries to resolve how teamwork and individual attributes are perceived by developers collaborating on GITHUB, and how those perceptions influence their work. Our findings can be used as complementary data to quantitative studies of developer behavior on GITHUB.

I. INTRODUCTION

Software development is technical and knowledge-intensive, but also human-centric and collaborative, benefiting from the social attributes of the people involved. Open Source Software (OSS) communities, in particular, tend to be quite diverse, with contributors ranging from professional developers to volunteers, all with varied personalities, educational and cultural backgrounds, age, gender, and expertise. Yet, despite participating in a very decentralized process, and despite this diversity, OSS teams often succeed to work together effectively and productively [1], [2].



Gender differences and bias in open source: pull request acceptance of women versus men

Josh Terrell¹, Andrew Kofink², Justin Middleton², Clarissa Rainear², Emerson Murphy-Hill², Chris Parnin² and Jon Stallings³

¹ Department of Computer Science, California Polytechnic State University—San Luis Obispo, San Luis Obispo, CA, United States

² Department of Computer Science, North Carolina State University, Raleigh, NC, United States

³ Department of Statistics, North Carolina State University, Raleigh, NC, United States

ABSTRACT

Biases against women in the workplace have been documented in a variety of studies. This paper presents a large scale study on gender bias, where we compare acceptance rates of contributions from men versus women in an open source software community. Surprisingly, our results show that women's contributions tend to be accepted more often than men's. However, for contributors who are outsiders to a project and their gender is identifiable, men's acceptance rates are higher. Our results suggest that although women on GitHub may be more competent overall, bias against them exists nonetheless.

OSS teams are typically more fluid and less tangible than their offline counterparts. They tend to form and dissolve organically around the task at hand, facing high turnover [17], while interactions between members are often limited to online channels [18]. In addition, GITHUB's implementation of the pull-based development model [19] enables anyone to submit changes to any repository with minimal effort, through pull requests (the so-called “drive-by” commits [13]). We wish to understand whether this unprecedented low barrier to entry for

Still,

Signals could help

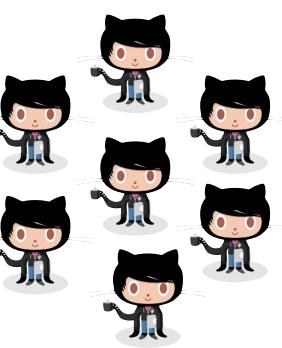
- disincentivize bad behavior
- match people to suitable mentors
- match people to suitable projects
- ...

Open source needs a **steady supply of effort by contributors**

But that is **harder today than ever before**
... because of how open source has **changed**

Low demographic diversity

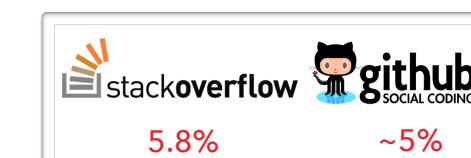
- Expectation
- Gender representation reality



"More about the contributions to the code than the 'characteristics' of the person"

"Any demographic identity is irrelevant"

"Code sees no color or gender"



In summary:
Many possible interventions

Missing: THEORY

- When and where to apply which intervention?
- What effects to expect?

Still,
Signals could help

- disincentivize bad behavior
- match people to suitable projects
- match people to suitable projects
- ...

Example #1 conclusion:
Ecosystem-level factors play an important role

New **signals** to display these otherwise unobservable ecosystem-level qualities:

- position in the network
- level of organizational support

The River of CPAN
Mon 20 April 2015

This blog post describes dependencies and reverse dependencies in CPAN. The river is Perl itself with all distributions in it. It contains all distributions upstream and downstream. Distributions sit somewhere along the river. Some distributions have reverse dependencies. That's what "upstream" means.

If you pollute a river you might cause problems for everyone downstream of you. And you're relying on the distributions upstream of you not polluting the river.

For CPAN, the pollution is bugs: if one of your upstream dists has a buggy version released to CPAN, it might break your distribution, but it might not. The further upstream a distribution, the more distributions it can potentially break, should it pollute the river.

So what?

CPAN authors / maintainers should know where their distributions sit on the river. We should help with that, and with visualising the upstream and downstream distributions. We should let authors know when a distribution moves up or down the river, particularly sudden large moves (if a distribution much further upstream starts using your distribution, you zoom to a position upstream of them).

<http://neilb.org/2015/04/20/river-of-cpan.html>

Example #2: "It's most important that the people seem nice"

How do people choose which project to contribute to?

Interviews:

15 GitHub users

Data:

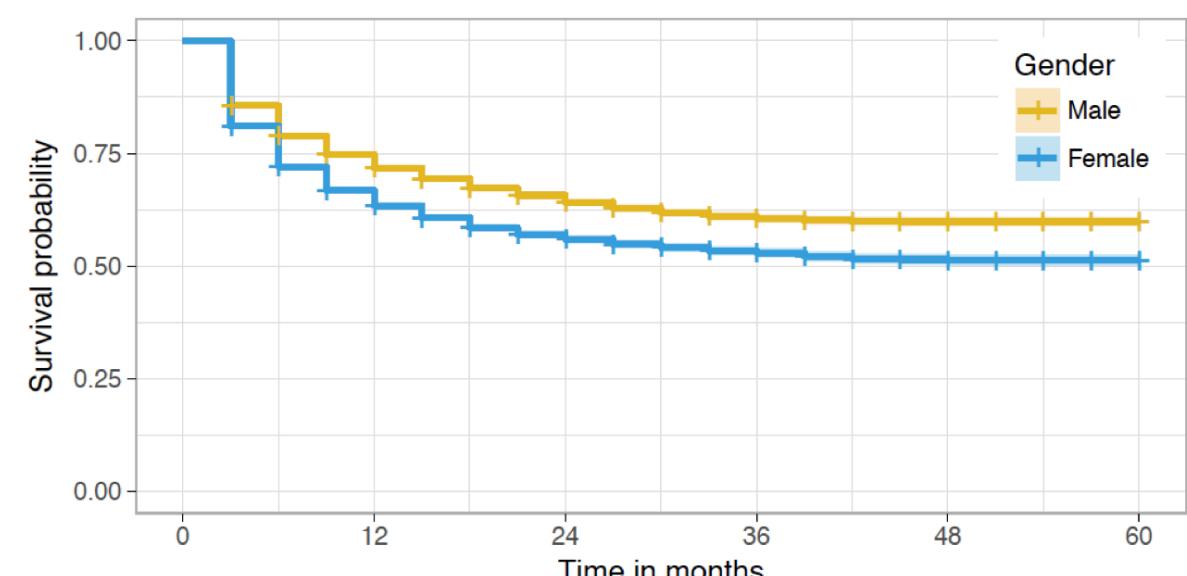
~10K npm packages

Model:

Logistic regression
(has new contributors)

The **tone of the community** is an important factor in both interviews and model.

Example #3: Building social capital
Why do women disengage earlier than men?



• Going Farther Together: The Impact of Social Capital on Sustained Participation in Open Source.
Qiu, H.S., Nolte, A., Brown, A., Serebrenik, A., and Vasilescu, B. ICSE 2019

Carnegie Mellon University School of Computer Science STRUDEL