Software Developer Diversity and Inclusion (SDDI) Workshop
Google — June 6, 2019

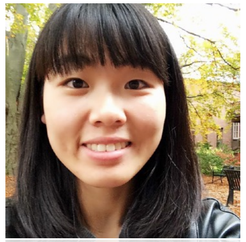
Leveraging Signals to Build More Sustainable Open Source Communities

Bogdan Vasilescu
@b_vasilescu

Acknowledgements


Courtney Miller


Anita Brown


Michelle Cao


Jim Herbsleb

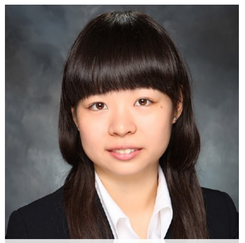

Christian Kästner


David Widder


Anita Sarma


Audris Mockus

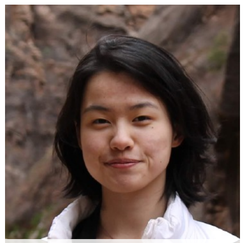

Alex Nolte


Sophie Qiu


Alex Serebrenik

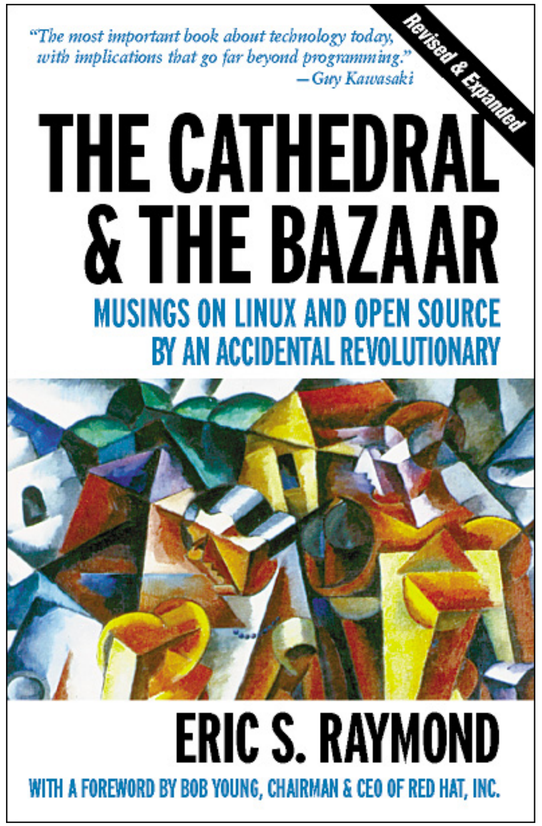

Marat Valiev

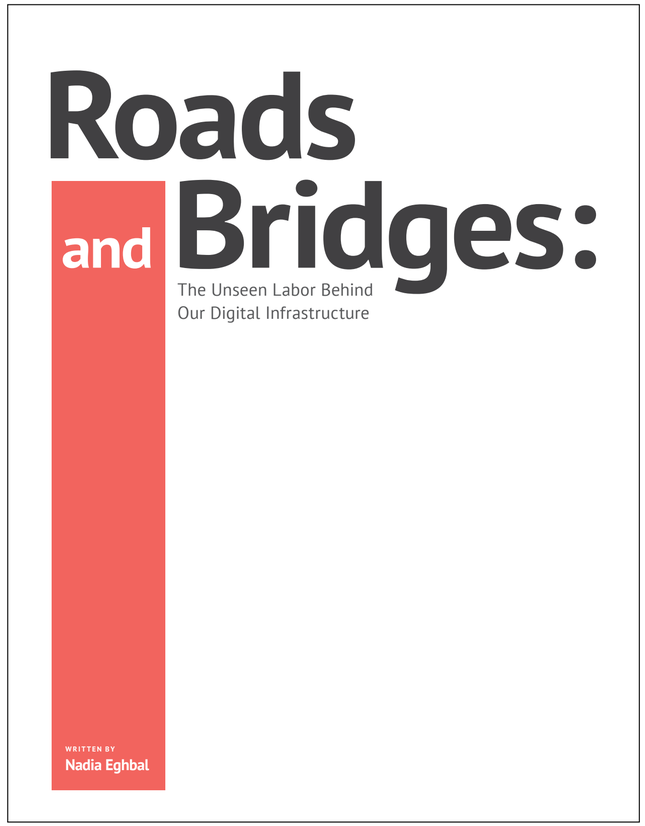

Laura Dabbish


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

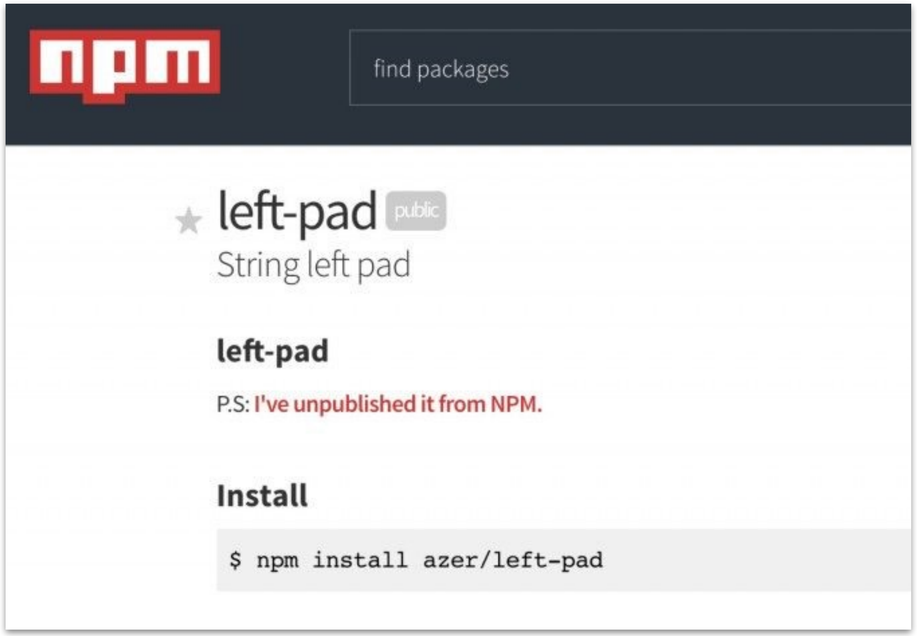
Economists: open source as “digital dark matter”

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 undermaintained 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

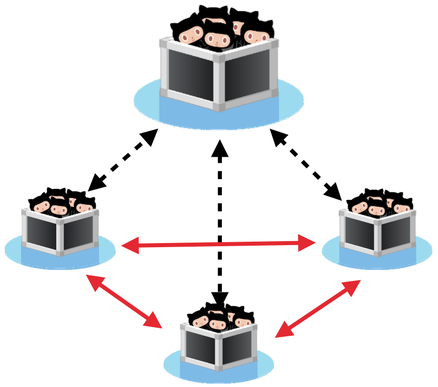
… because of how open source has changed

Today: more problems than solutions

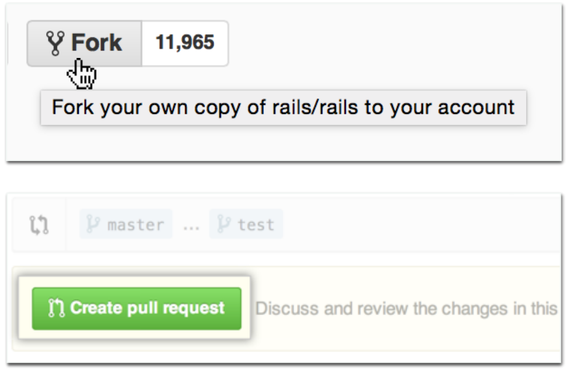
How has open source changed?

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

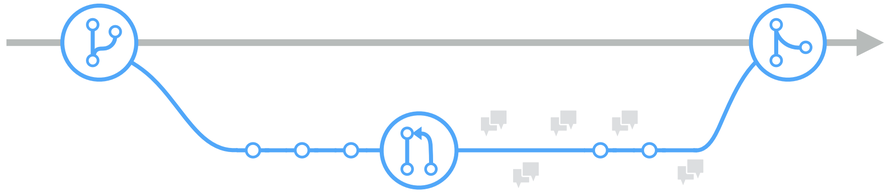
- Git version control



- 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

github SOCIAL CODING

100 million repositories

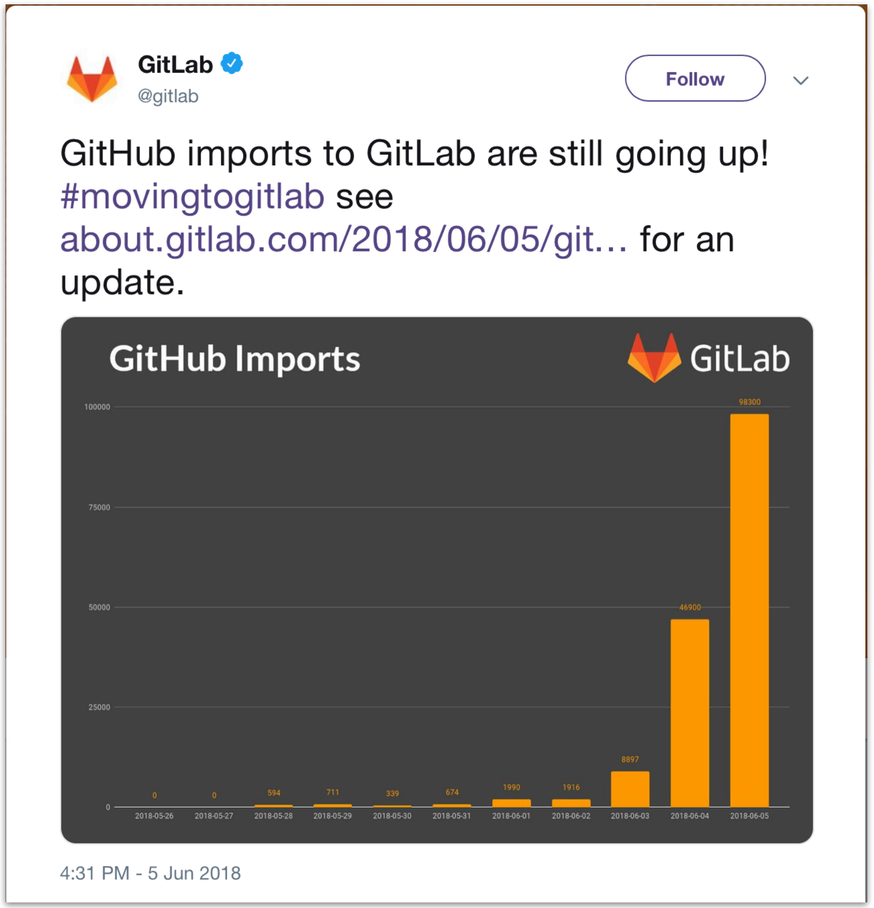
31 million users

(November 2018)

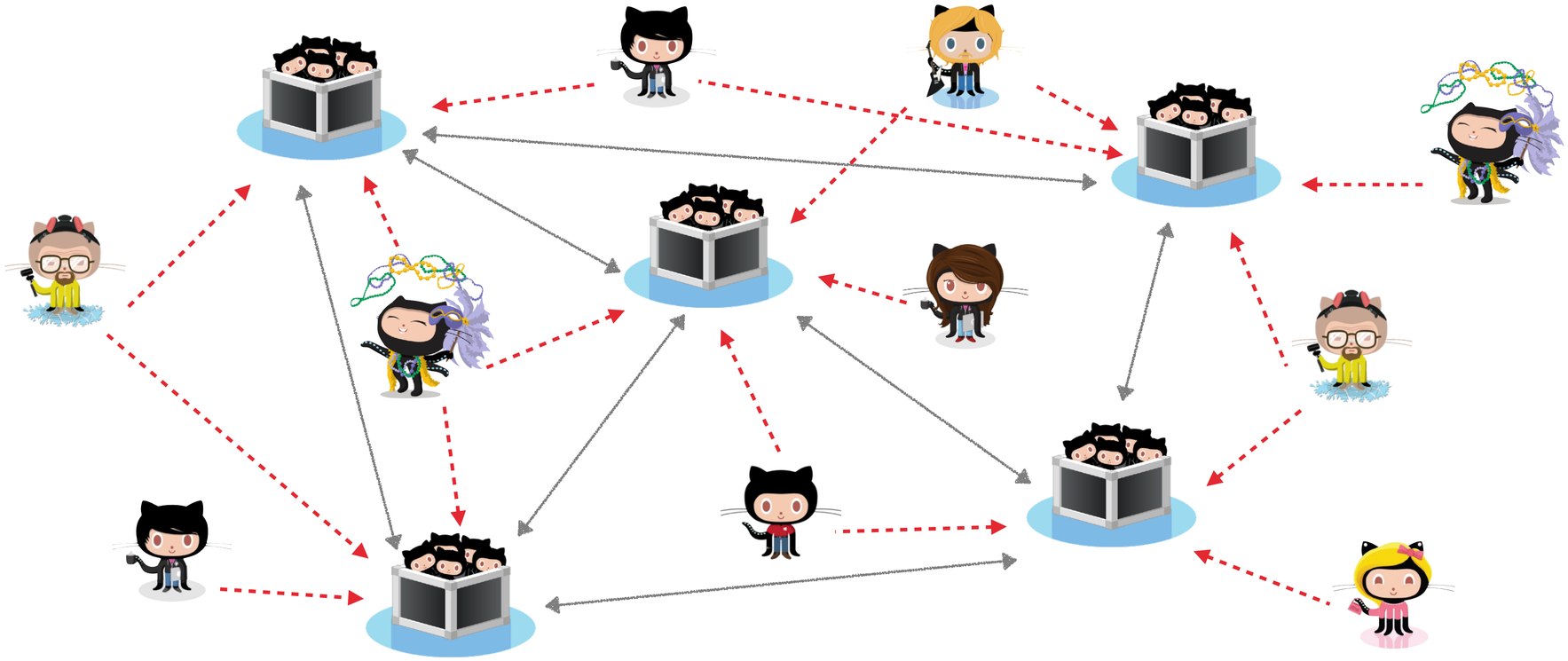
Bitbucket

6 million users

(March 2019)



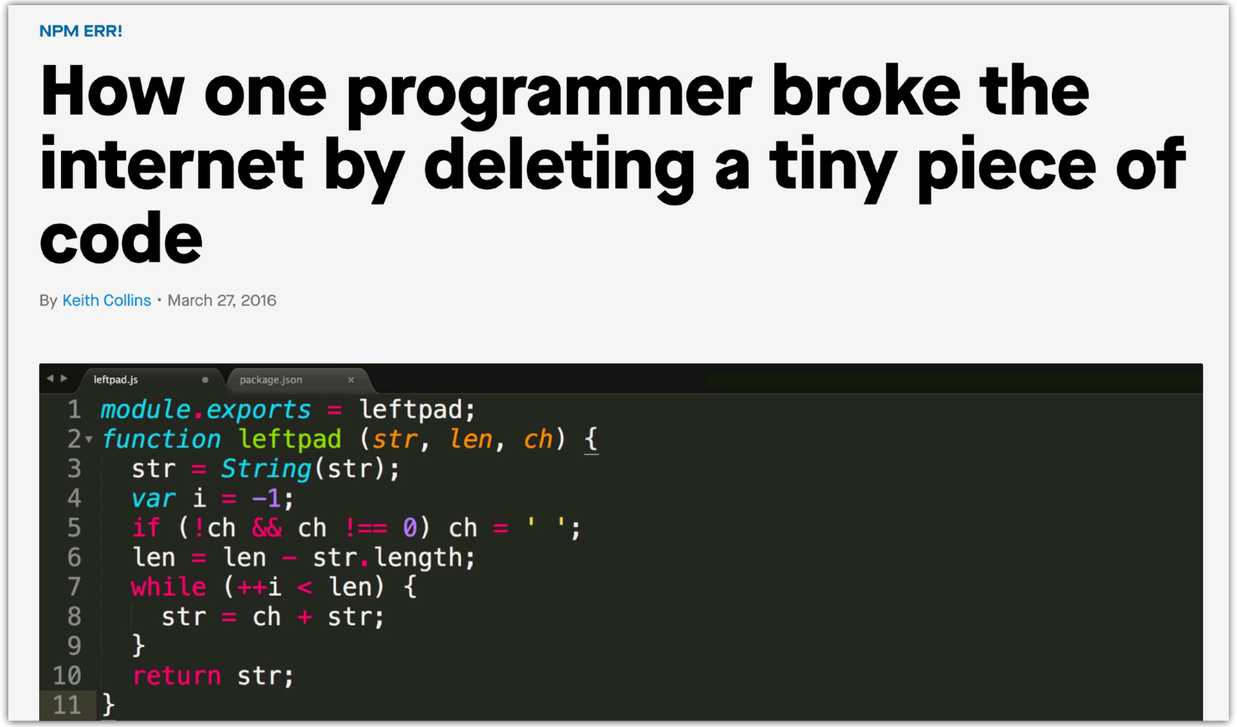
Change #2:
Complex *ecosystem* of interdependencies



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)
- …



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)

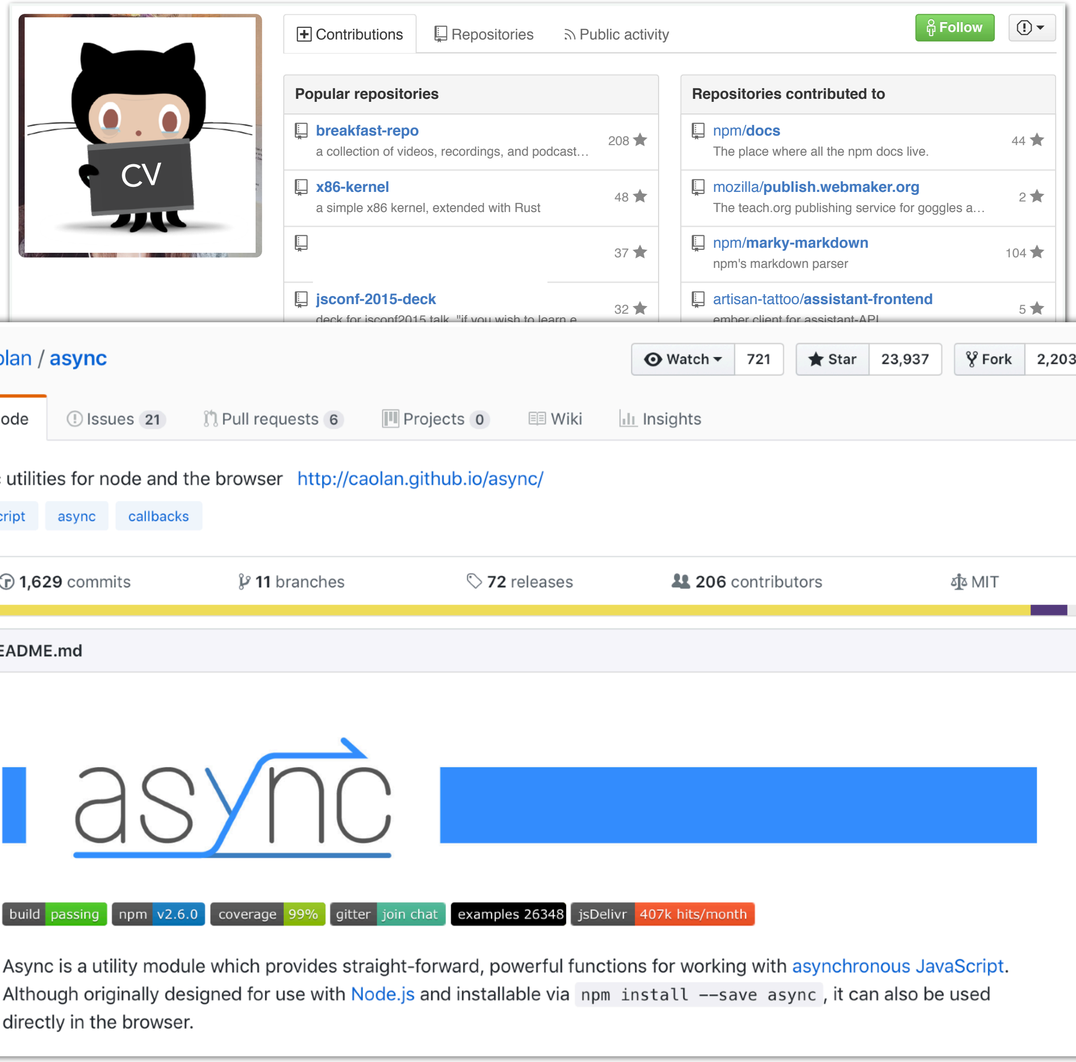
- Currently
  - Lots of commercial involvement
    - Companies (Go - Google, React - Facebook, Swift - Apple)
    - Startups (Docker, npm, Meteor)

- 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)



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

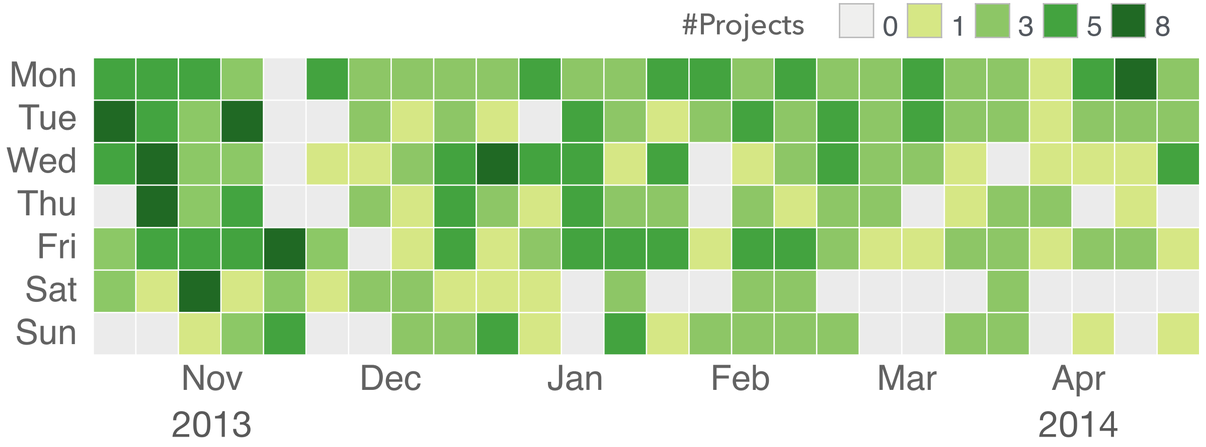
https://www.zdnet.com/article/equifax-confirms-apache-struts-flaw-it-failed-to-patch-was-to-blame-for-data-breach/

High level of demands & stress

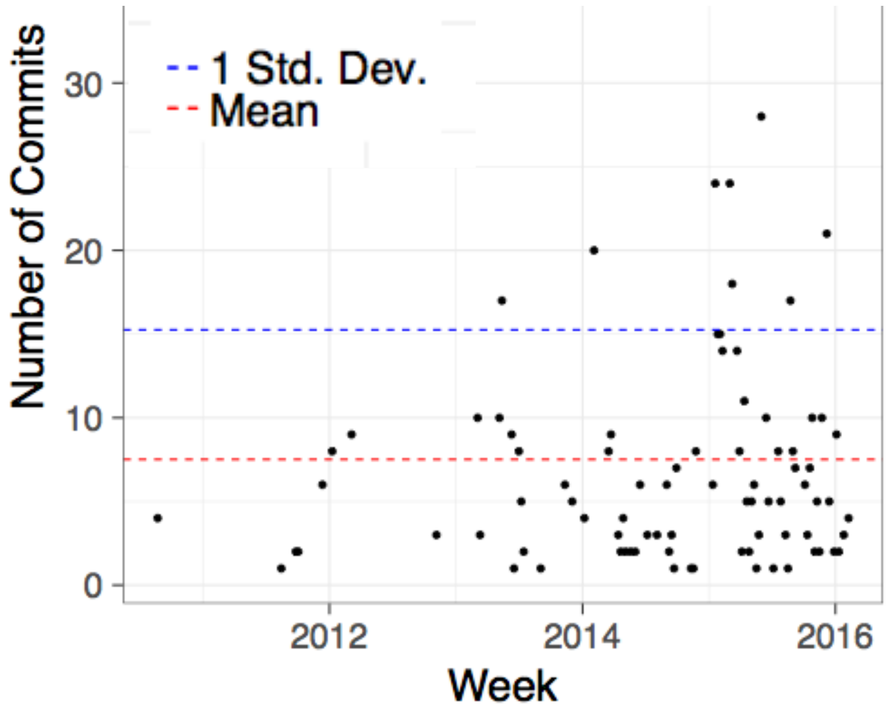
- Easy to report issues / submit PRs
  - Growing volume of requests
- Social pressure to respond quickly
  - Otherwise, off-putting to newcomers (Steinmacher et al. 2015)
- Entitlement, unreasonable requests from users:
  - *“I have been waiting 2 years for Angular to track the ‘progress’ event and it 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 it's 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/

https://github.com/isaacs/github/issues/627

Low demographic diversity

- Gender representation reality

stackoverflow 5.8%

github SOCIAL CODING ~5%

open source 10.9%

Google 18%

Microsoft 16.6%

- 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*"

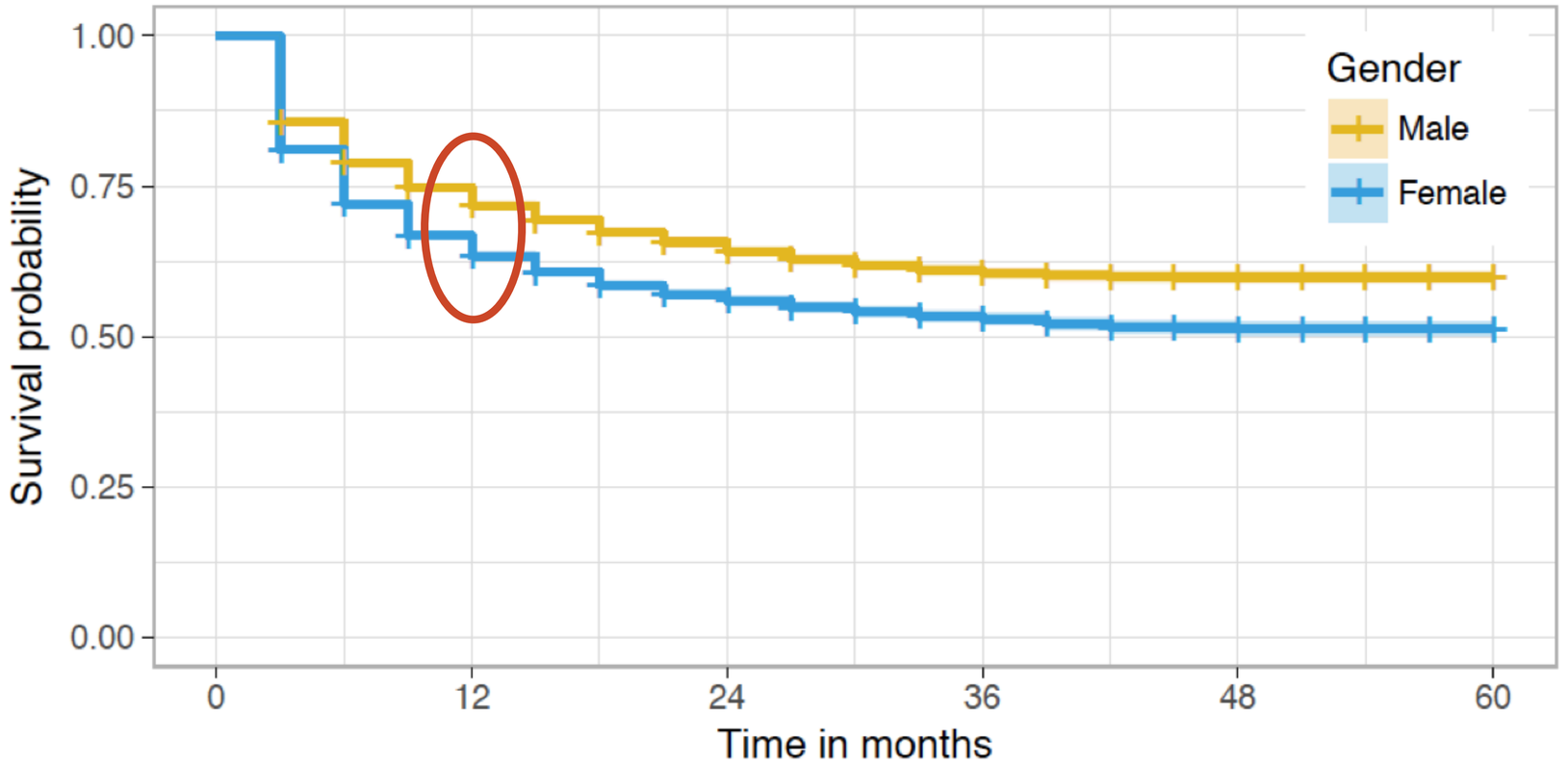
- FLOSS 2013: A survey dataset about free software contributors: challenges for curating, sharing, and combining G Robles, L Arjona-Reina, B Vasilescu, A Serebrenik, JM Gonzalez-Barahona. *MSR 2014*
- Google Diversity (2015) www.google.com/diversity/index.html#chart
- Inside Microsoft (2015) https://goo.gl/nT4YiI

- 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*

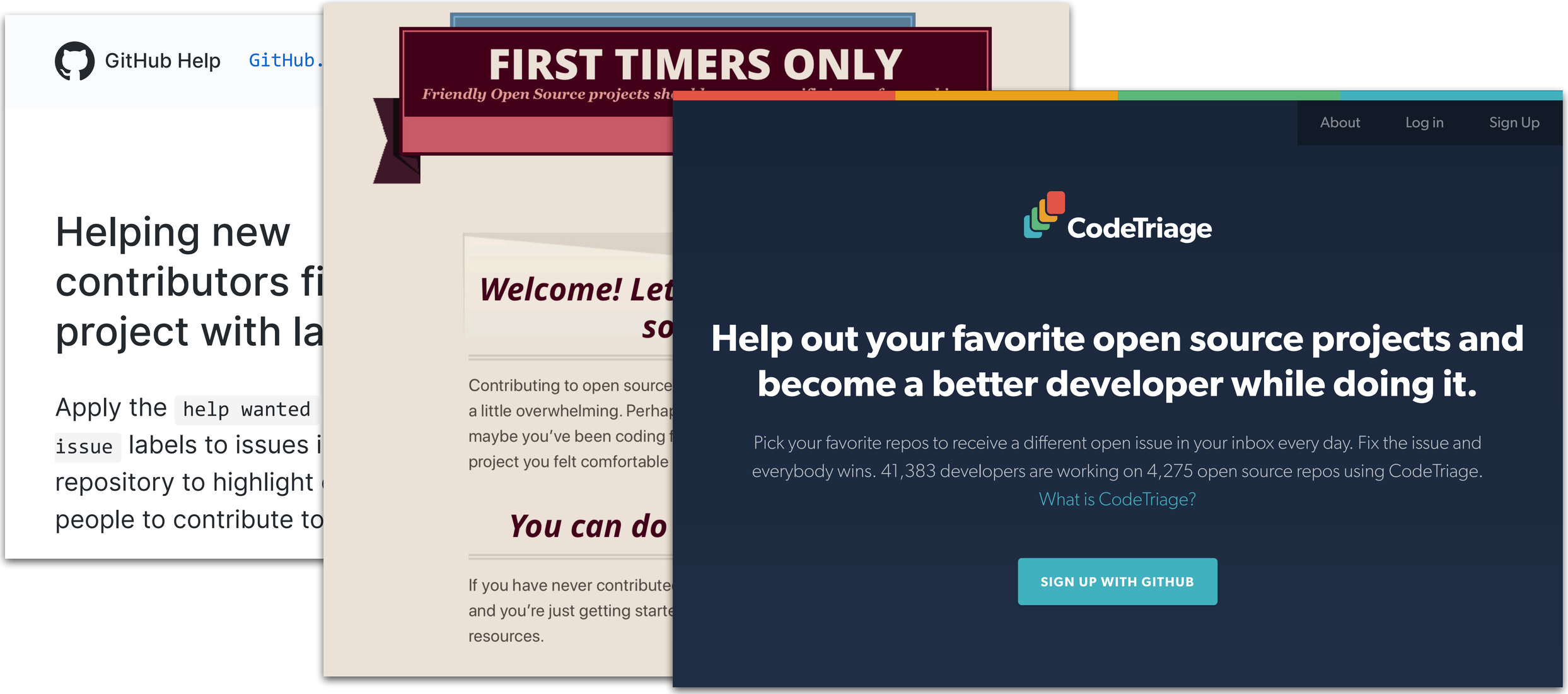


*Among committers with first & last names on their profiles

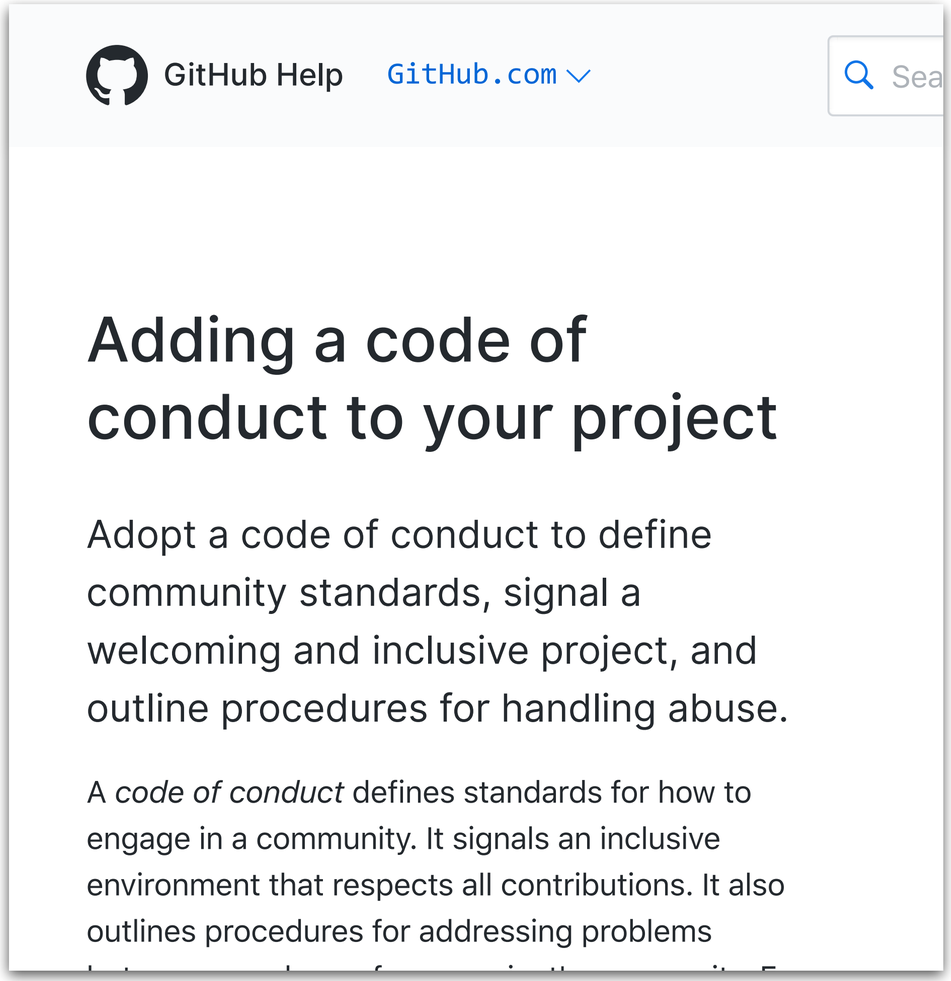
- Going Farther Together: The Impact of Social Capital on Sustained Participation in Open Source. Qiu, H.S., Nolte, A., Brown, A., Serebrenik, A., and Vasilescu, B. *ICSE 2019*

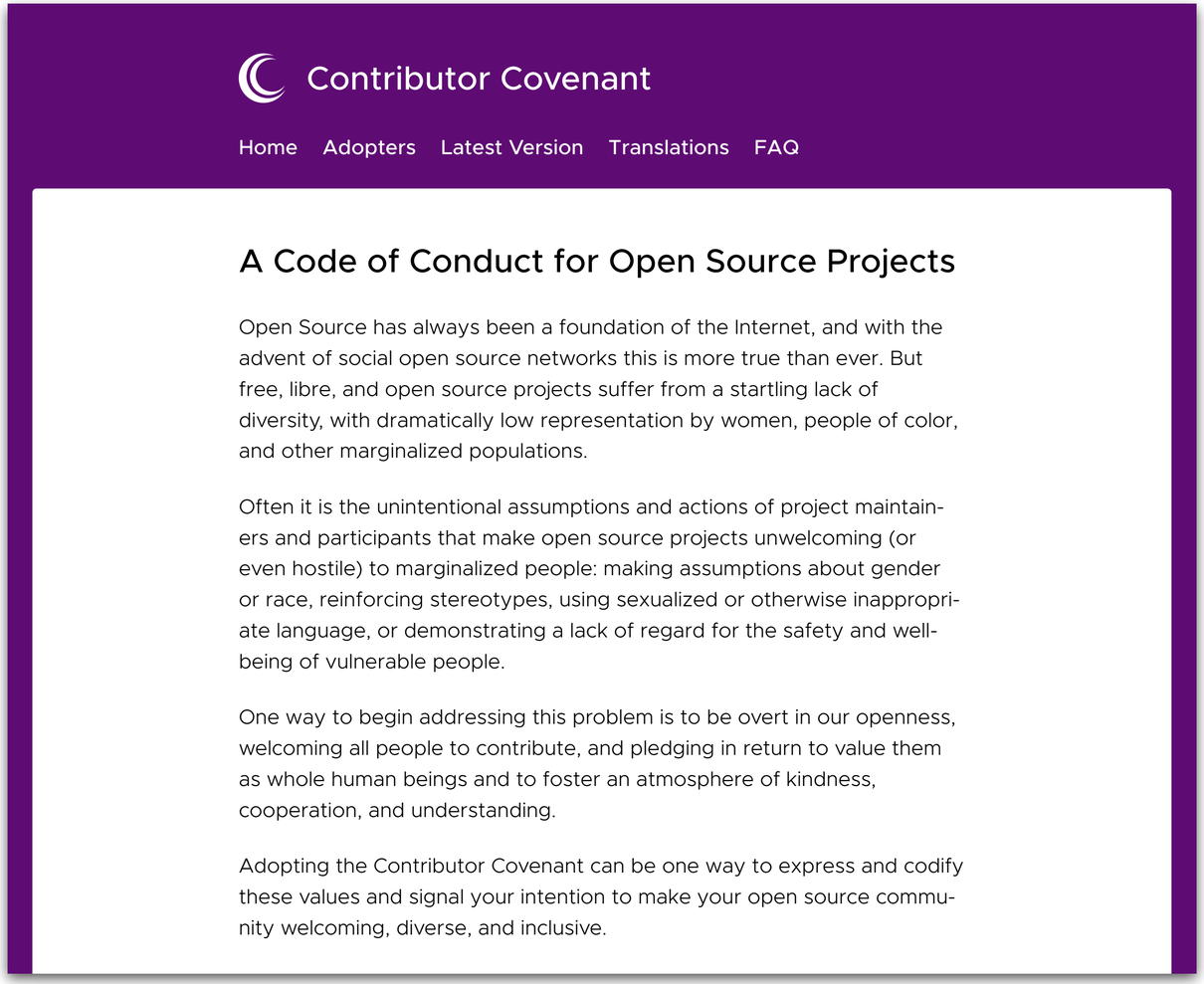
What are people doing to attract & retain (diverse) contributors to open source?

Recent trend: Guides for newcomers



Recent trend: Codes of conduct





(Tourani, Adams, & Serebrenik, SANER 2017)

Recent trend: Safe spaces

Table 1. OSS websites with women only spaces.

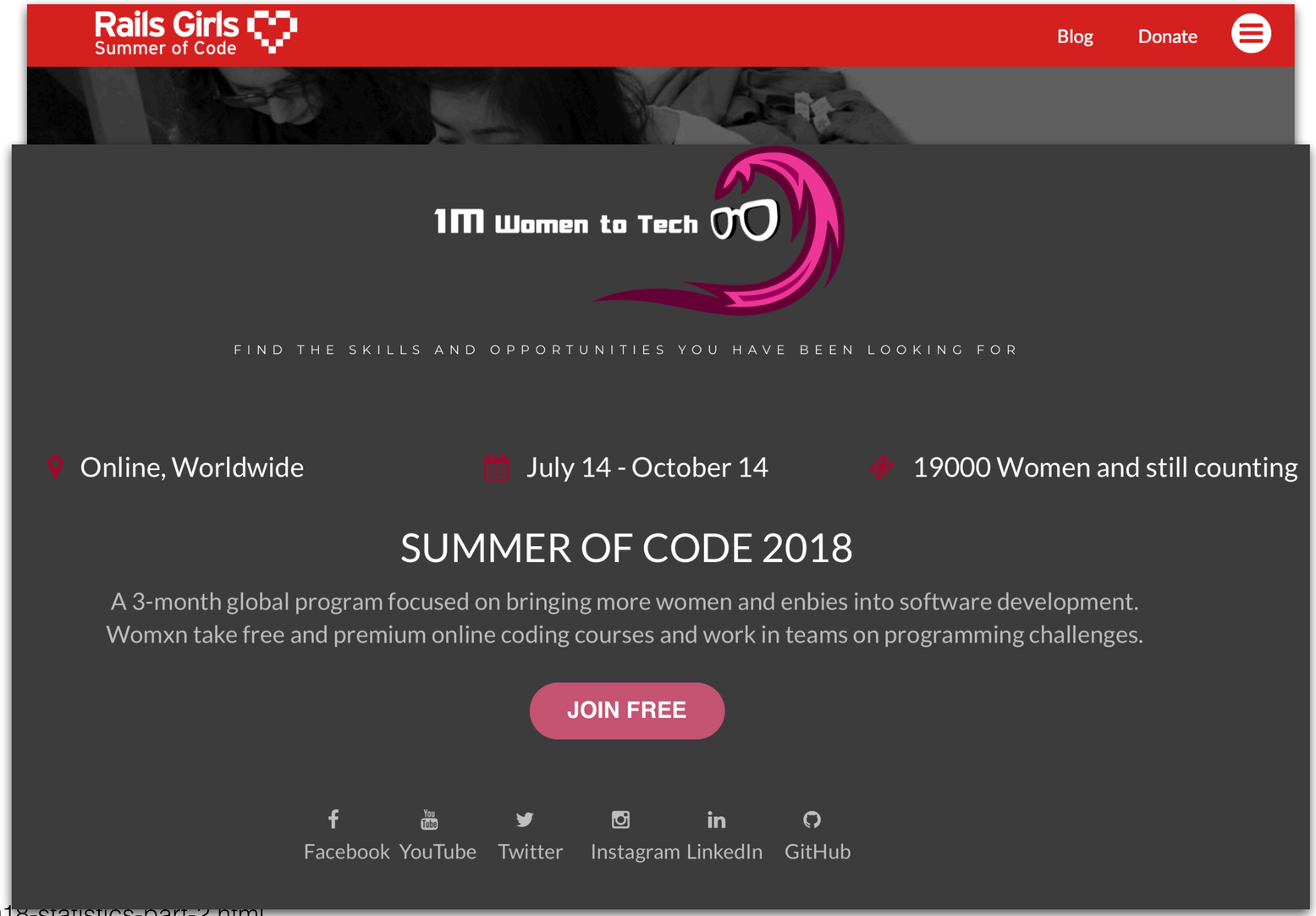
Software package	Name of the space	URL
ArchLinux	Arch Linux for Women	http://archwomen.org
Bitcoin	Women in Bitcoin Madchenabend in Berlin	https://www.facebook.com/womeninbitcoin/
BonitaSoft	Blog Post about Community efforts for encouraging women	https://community.bonitasoft.com/behind-scenes-bonita-21-27-feb-2011
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

Google Summer of Code

2018: “11.63% of accepted students are women”



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

Recent trend: Hackathons

ANITA B.ORG

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

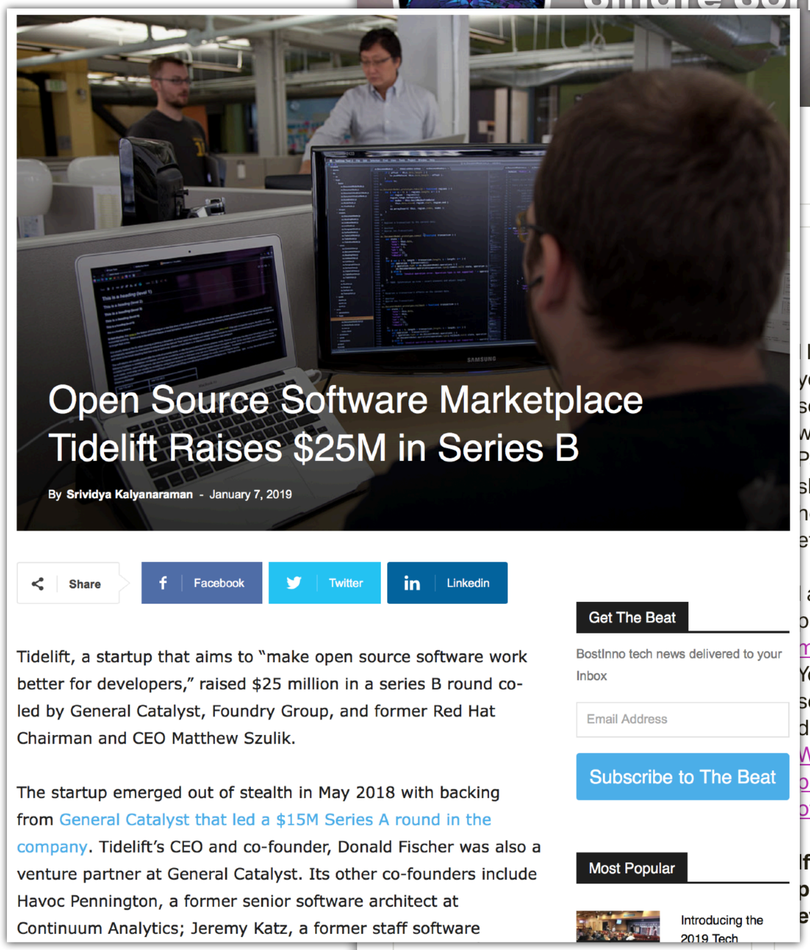
Grace Hopper Celebration | Register | Awards | Speakers | Sponsors | Acade

2018 Open Source Day

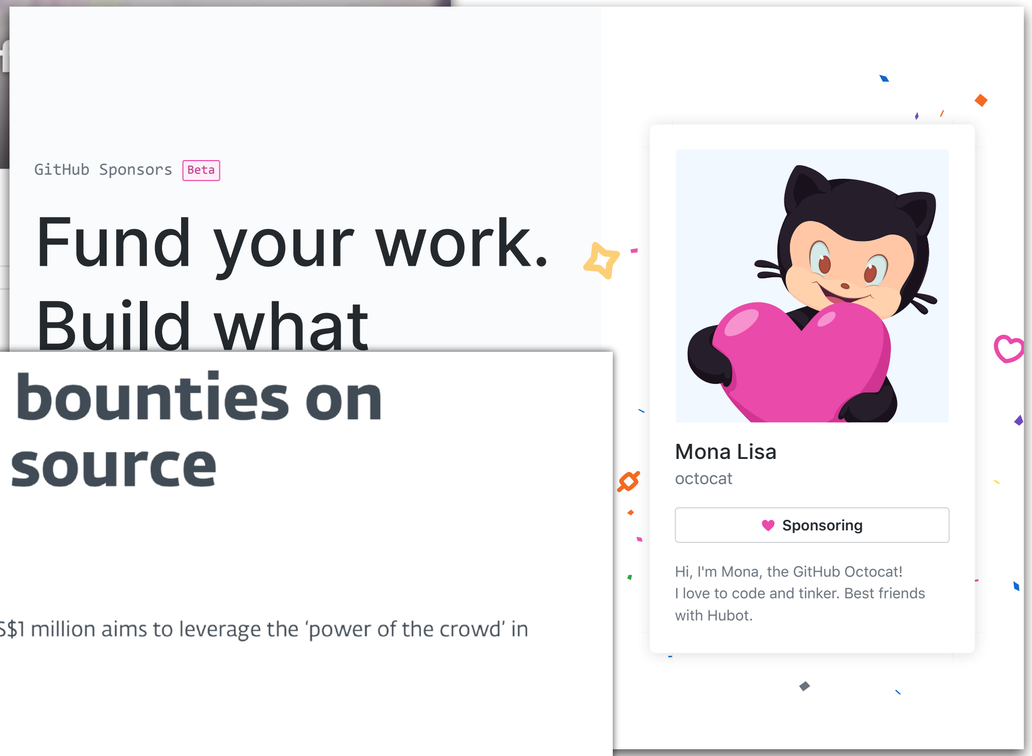
Thursday, September 27 | 9:30 a.m. – 5:30 p.m.

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.

Recent trend: New forms of funding



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



https://github.com/sponsors

https://www.welivesecurity.com/2019/01/07/eu-bounty-bugs-open-source-software/

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

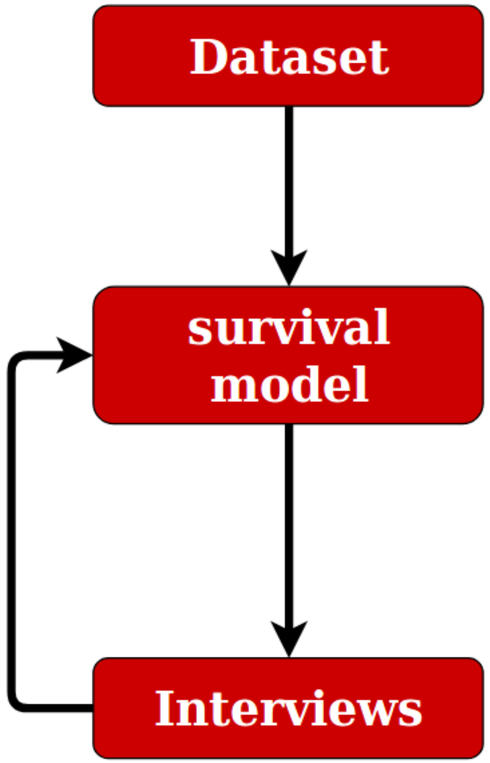
https://zenodo.org/record/1297925

Model:

Cox survival regression

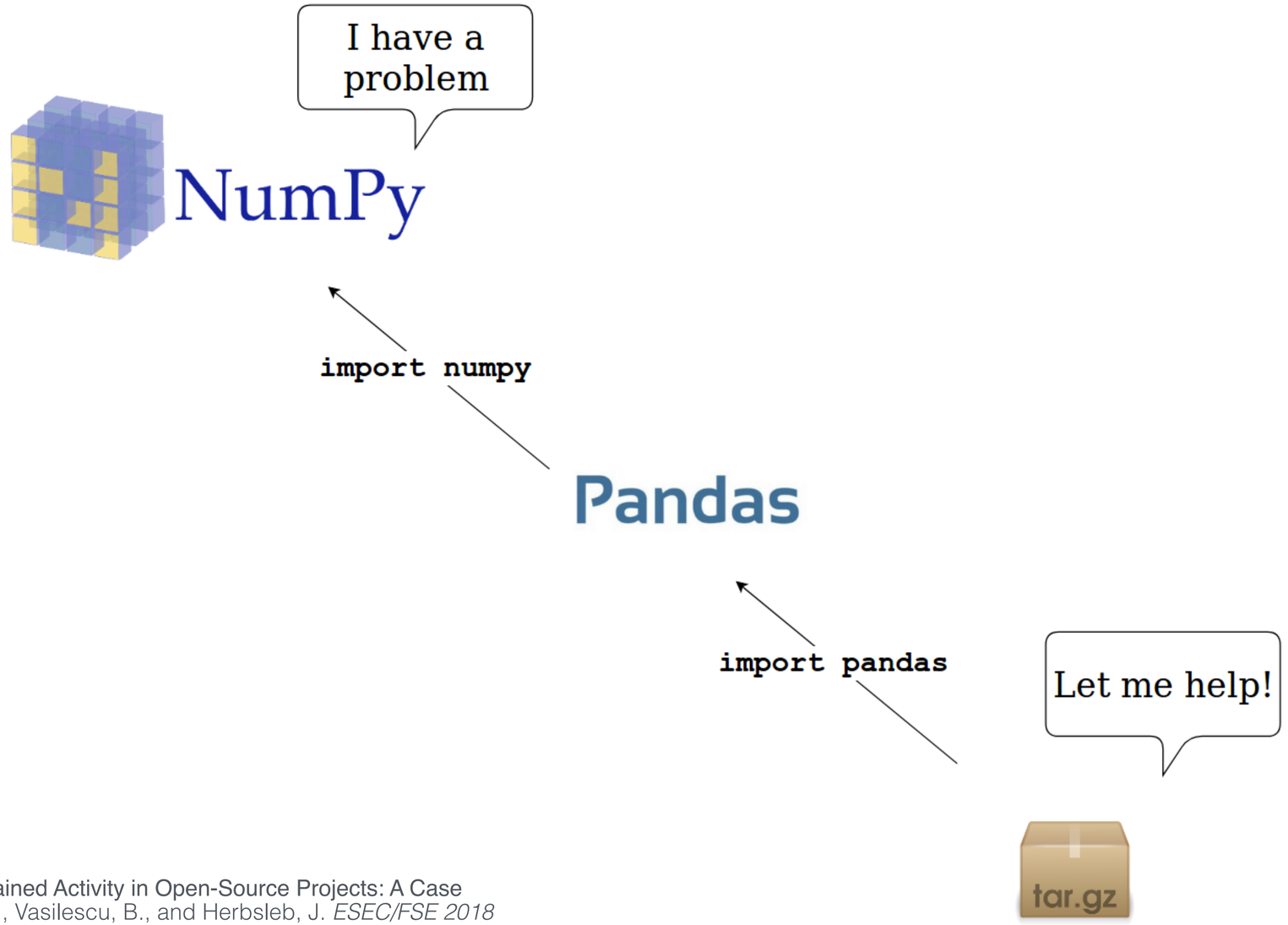
Interviews:

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*

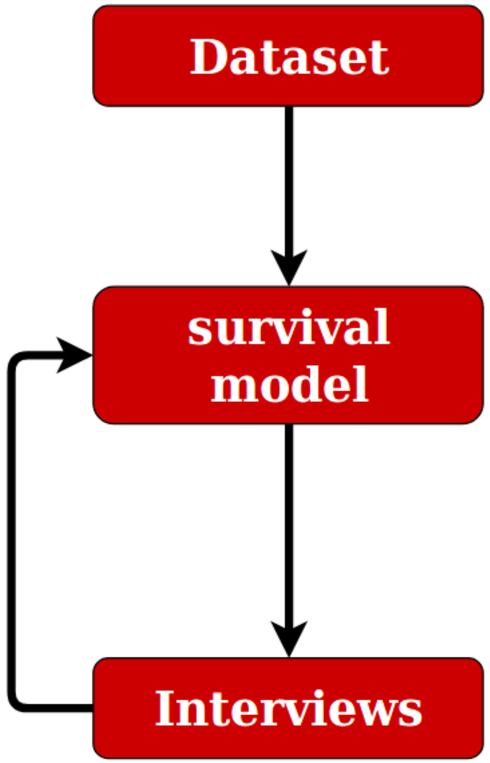
Transitive downstream dependencies are harmful

Feature: Katz centrality
(discounted transitive dependencies)

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

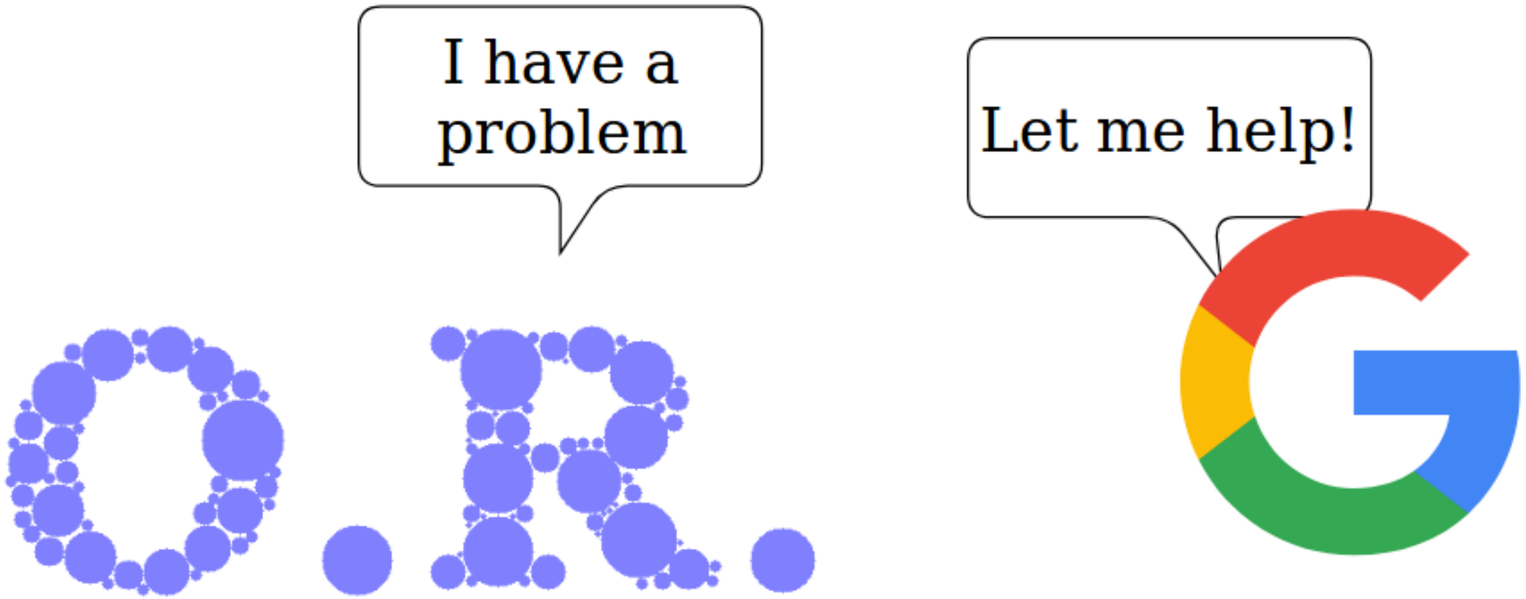
Interviews:

- 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:

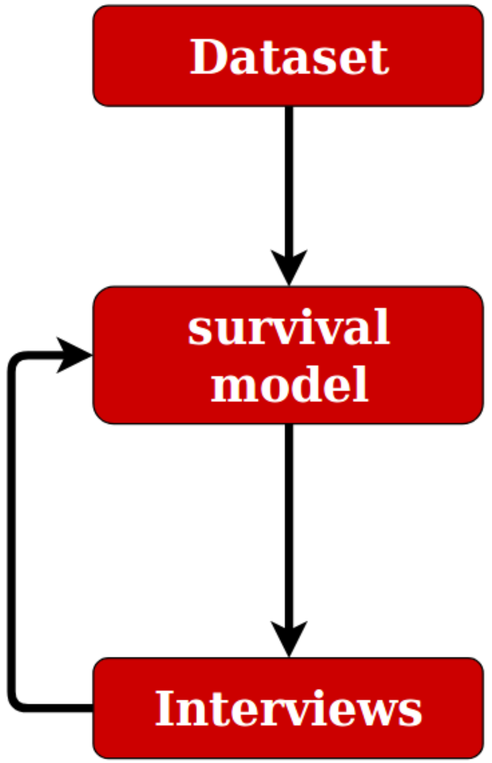
high commercial involvement

Early stage: **-51%** survival

Long term: **-15%**

Interviews:

- companies bring more resources
- but they can withdraw anytime

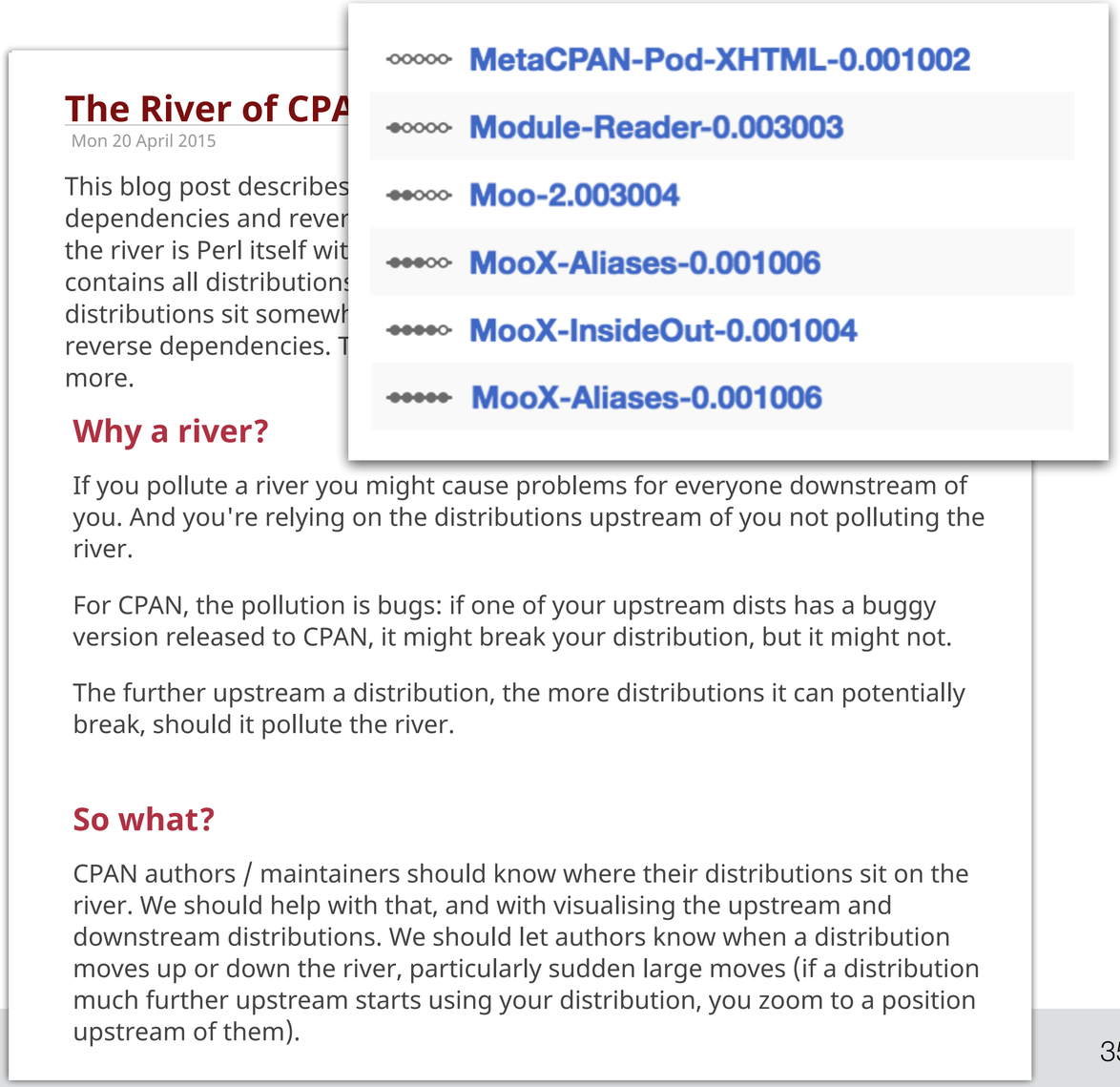


- Ecosystem-Level Determinants of Sustained Activity in Open-Source Projects: A Case Study of the PyPI Ecosystem. Valiev, M., Vasilescu, B., and Herbsleb, J. *ESEC/FSE 2018*

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



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)

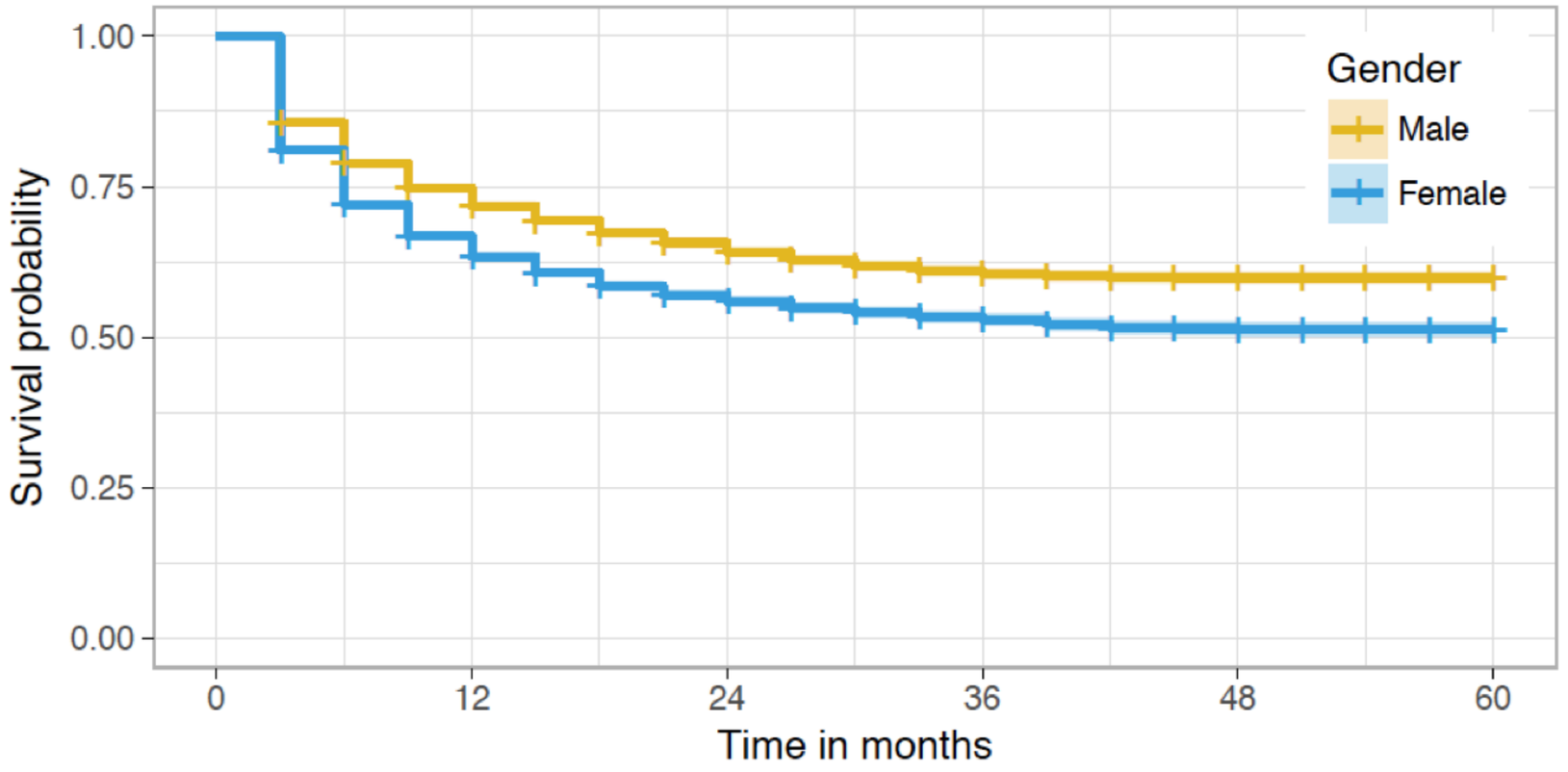
The tone of the community is an important factor in both interviews and model.

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

- 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*

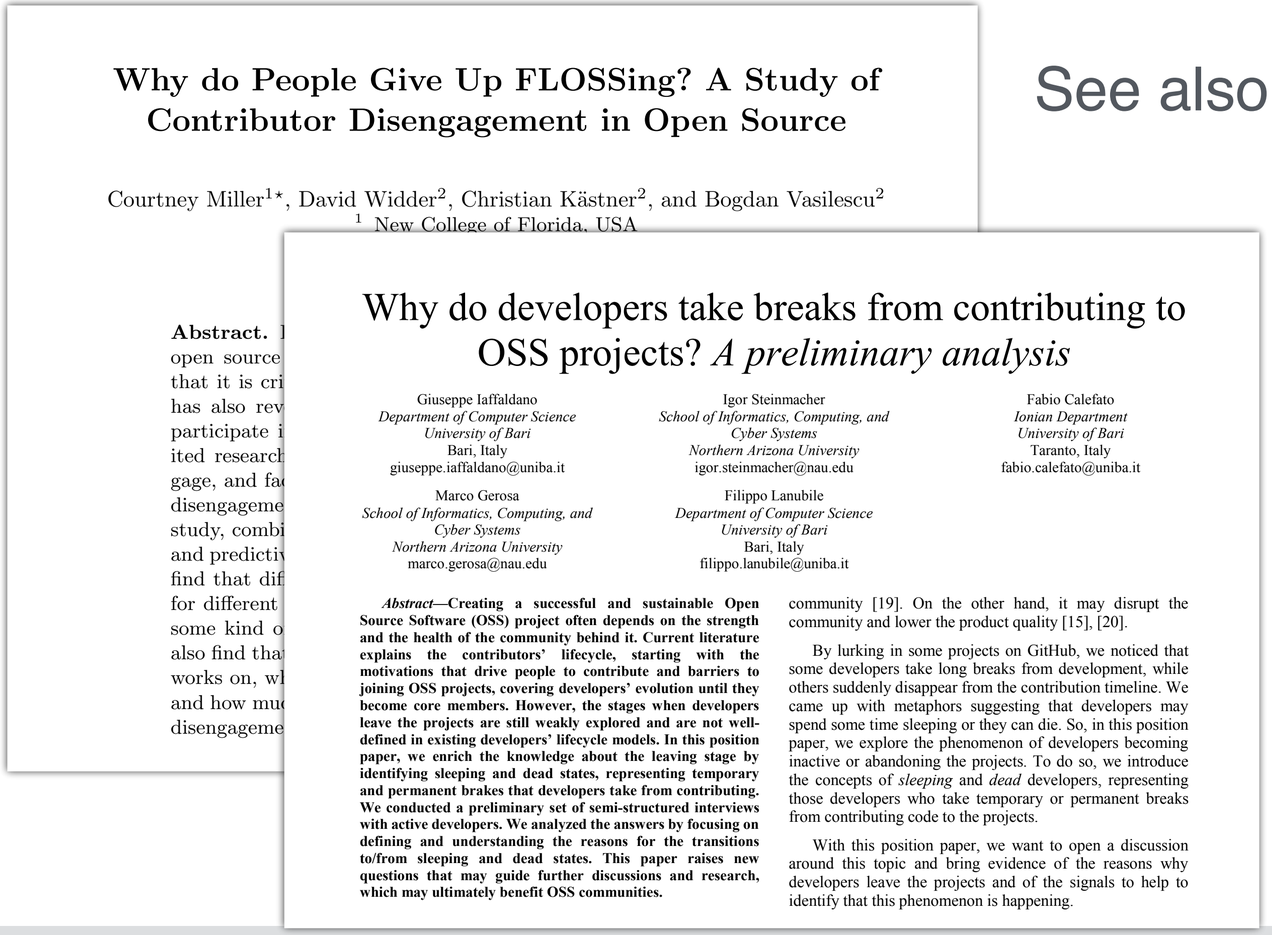
Example #3: Building social capital

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*

See also:



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

Courtney Miller[1]*, David Widder[2], Christian Kästner[2], and Bogdan Vasilescu[2]
[1] New College of Florida, USA

Abstract. I
open source
that it is cri
has also rev
participate i
ited research
gage, and fa
disengageme
study, combi
and predictiv
find that dif
for different
some kind o
also find tha
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

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

Fabio Calefato
Ionian Department
University of Bari
Taranto, Italy
fabio.calefato@uniba.it

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

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

***Abstract*—Creating a successful and sustainable Open Source Software (OSS) project often depends on the strength and the health of the community behind it. Current literature explains the contributors' lifecycle, starting with the motivations that drive people to contribute and barriers to joining OSS projects, covering developers' evolution until they become core members. However, the stages when developers leave the projects are still weakly explored and are not well-defined in existing developers' lifecycle models. In this position paper, we enrich the knowledge about the leaving stage by identifying sleeping and dead states, representing temporary and permanent brakes that developers take from contributing. We conducted a preliminary set of semi-structured interviews with active developers. We analyzed the answers by focusing on defining and understanding the reasons for the transitions to/from sleeping and dead states. This paper raises new questions that may guide further discussions and research, which may ultimately benefit OSS communities.**

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

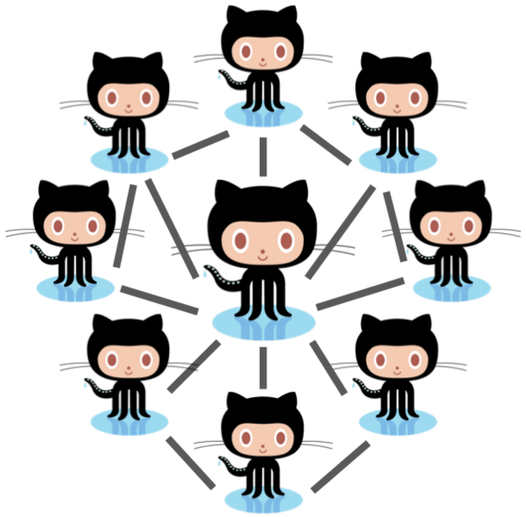
By lurking in some projects on GitHub, we noticed that some developers take long breaks from development, while others suddenly disappear from the contribution timeline. We came up with metaphors suggesting that developers may spend some time sleeping or they can die. So, in this position paper, we explore 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.

- 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*

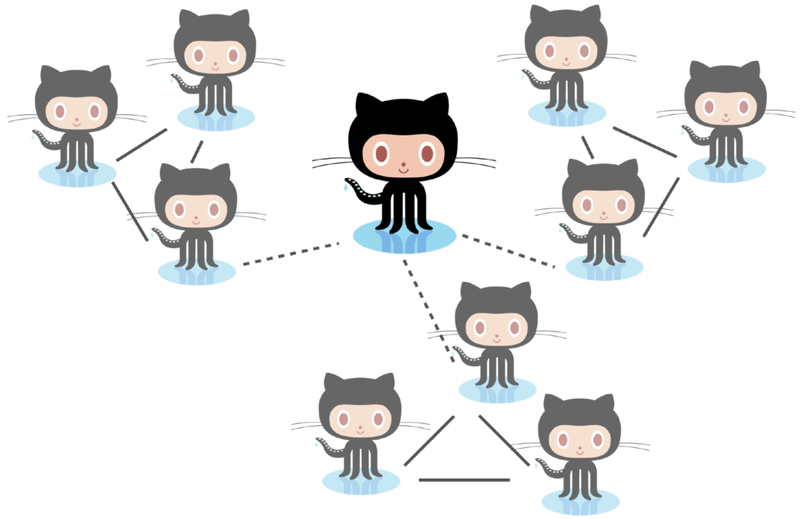
Social capital theory explains long-term engagement

Bonding social capital: benefiting from strongly connected network



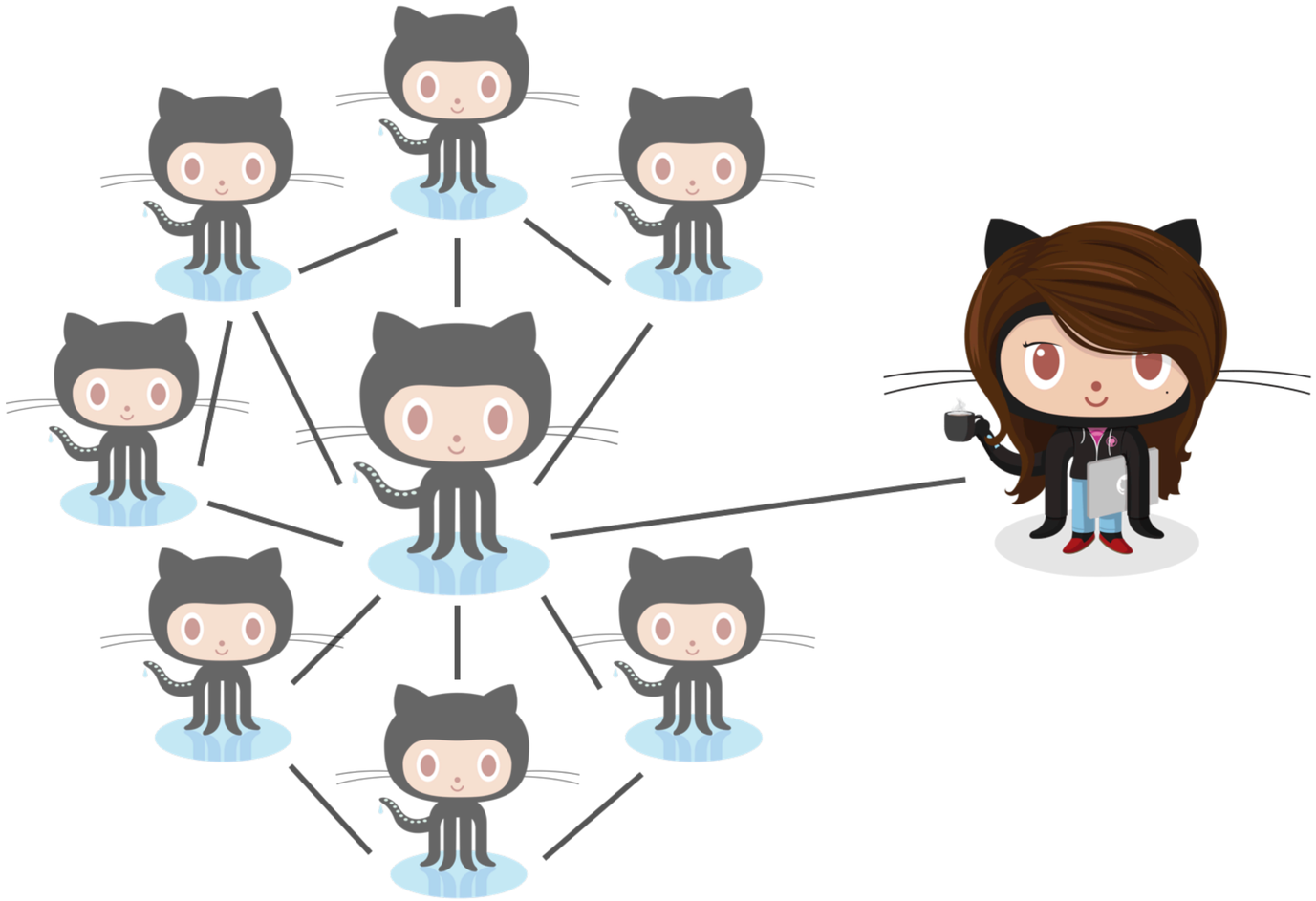
Willingness to continue
(Coleman, 1990)

Bridging social capital: benefiting from network with diverse info



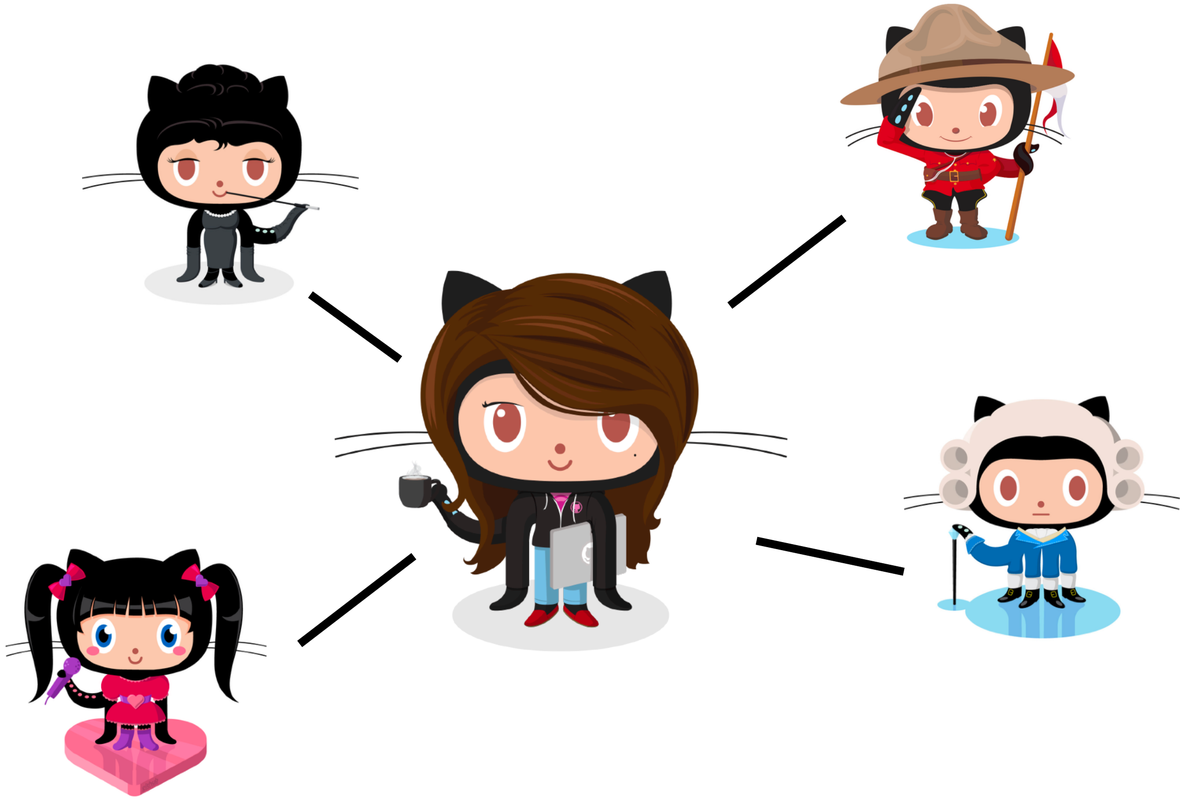
Opportunity to continue
(Burt, 1998, 2001)

Cohesive networks might foster discrimination / exclusion

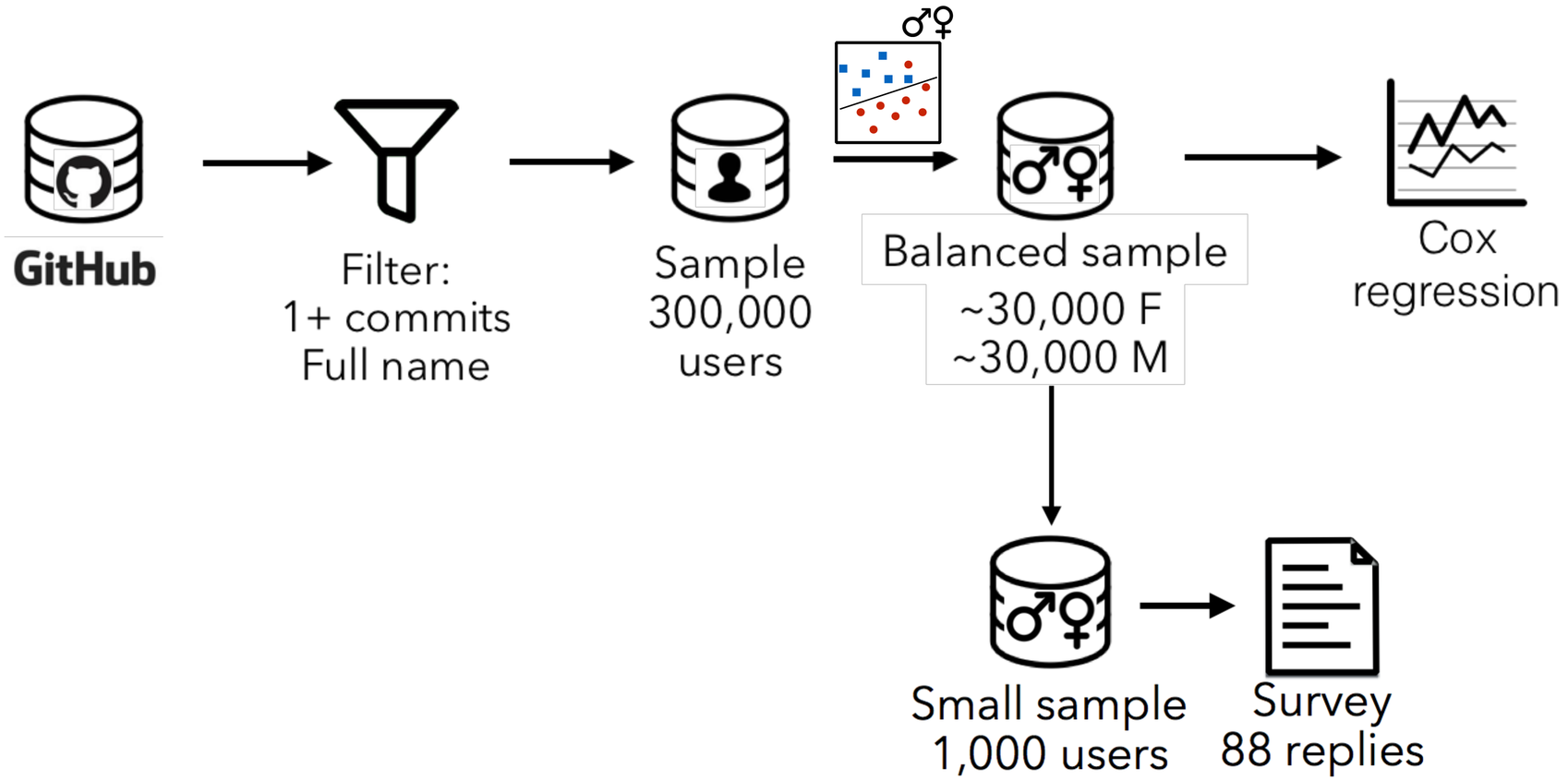


Being part of teams with more diverse information ~ more prolonged engagement, esp. for women

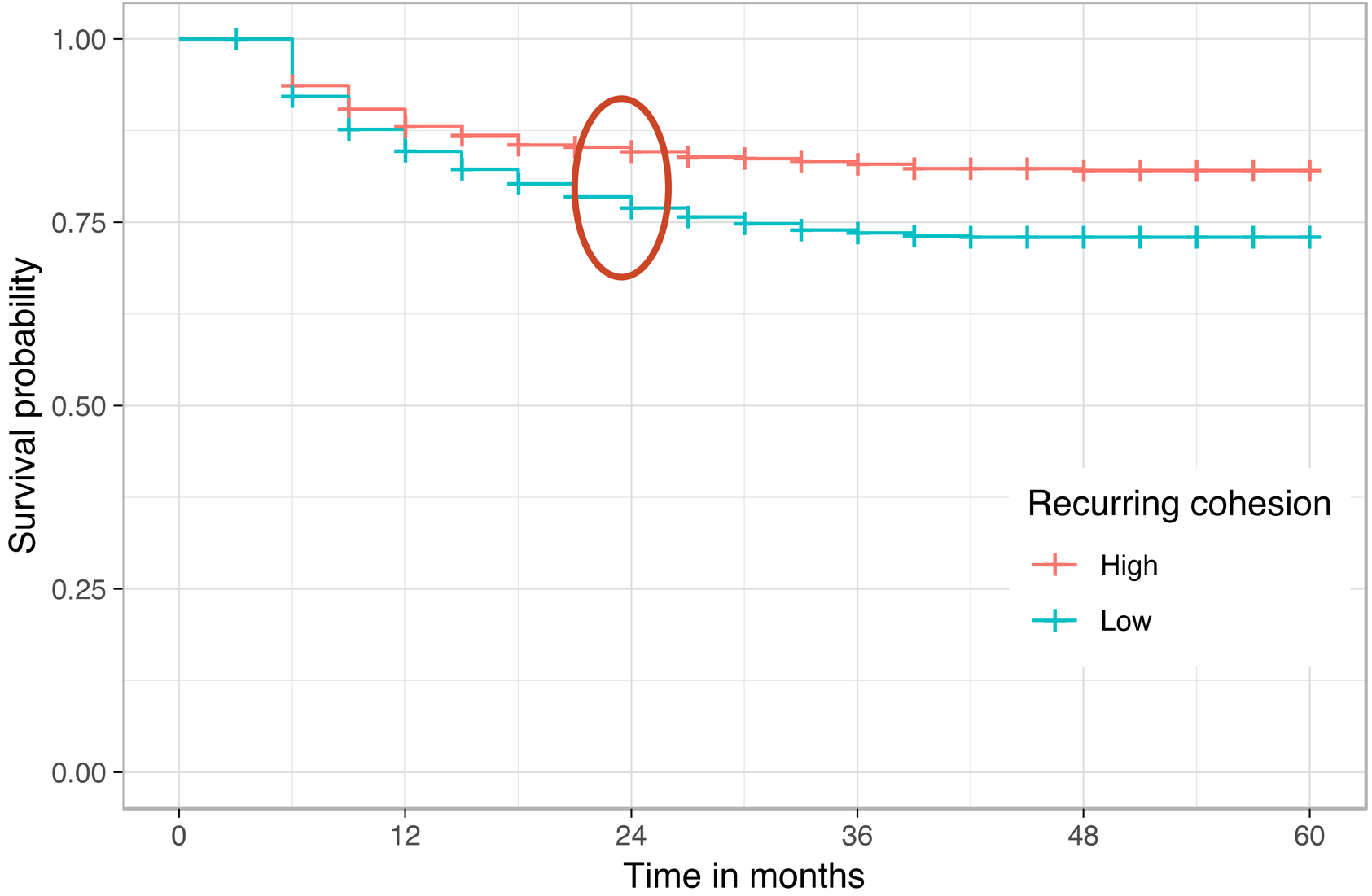
Information diversity should reduce the risk of demographic-based echo chambers.



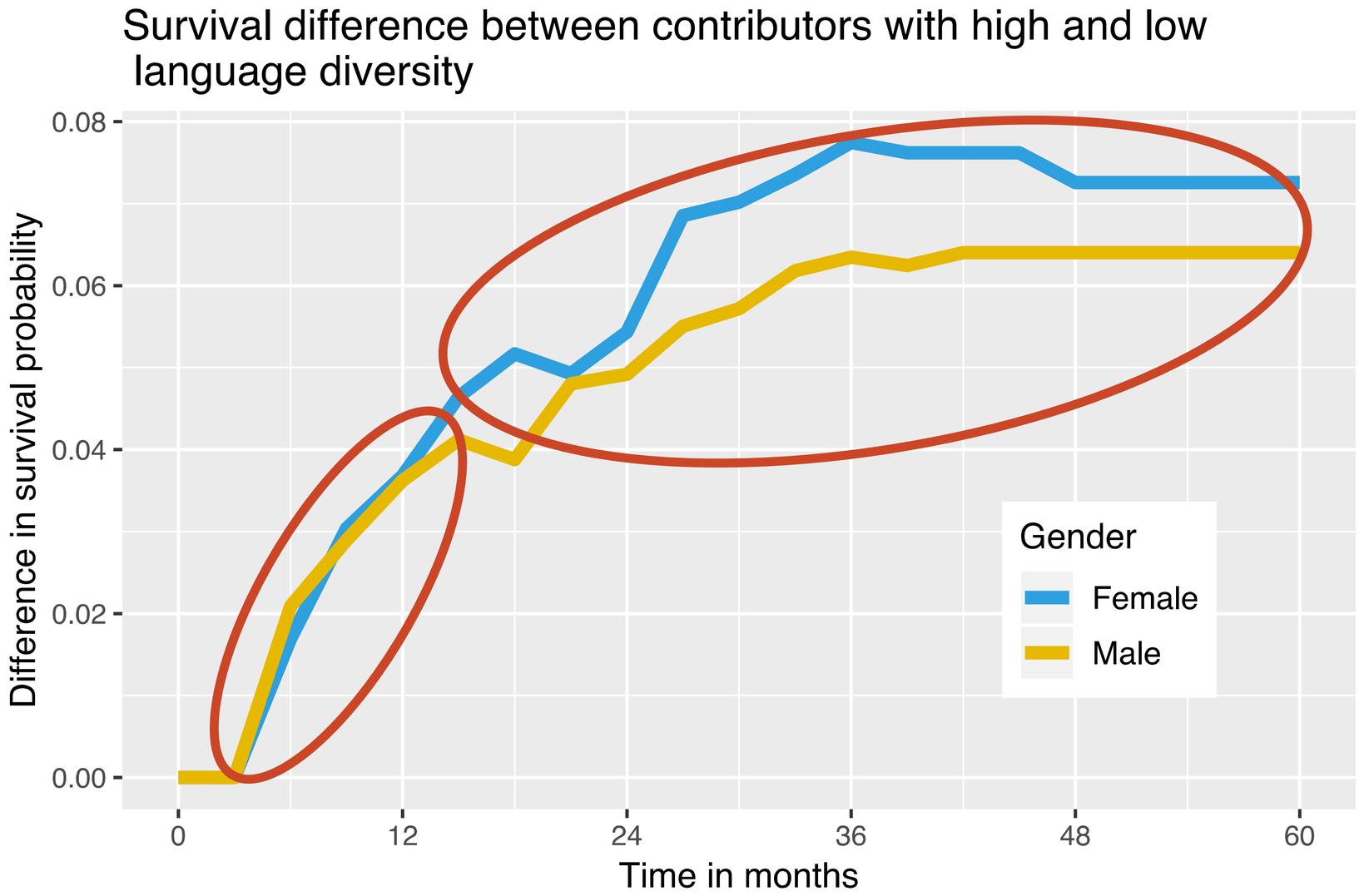
Large-scale mixed-methods study



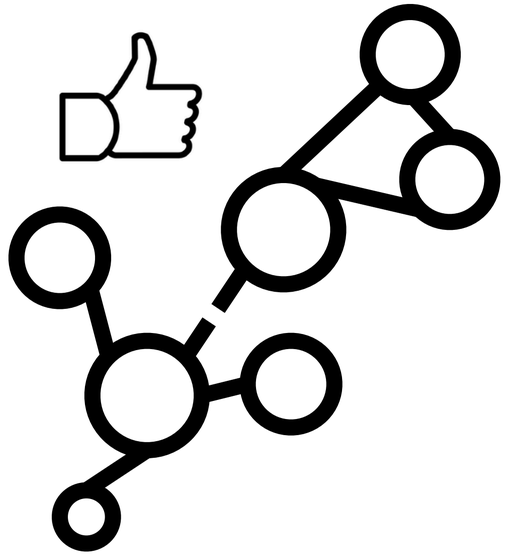
More social capital ~ more prolonged engagement



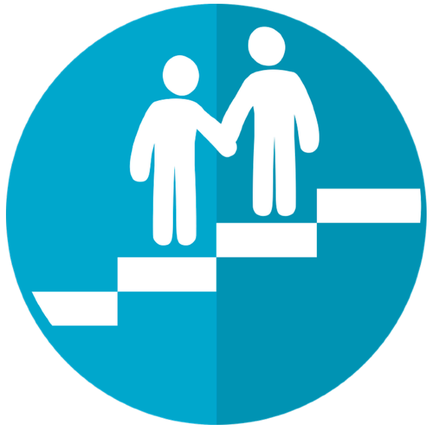
Women in language- (informationally-) diverse teams disengage at lower rates



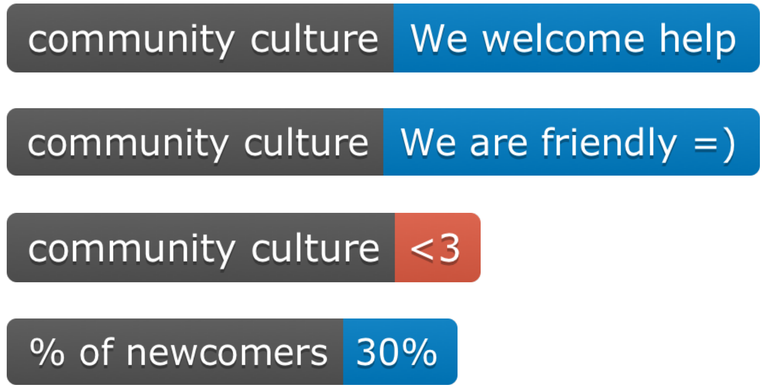
Example #3 conclusion:



Recommend projects that can help build social capital



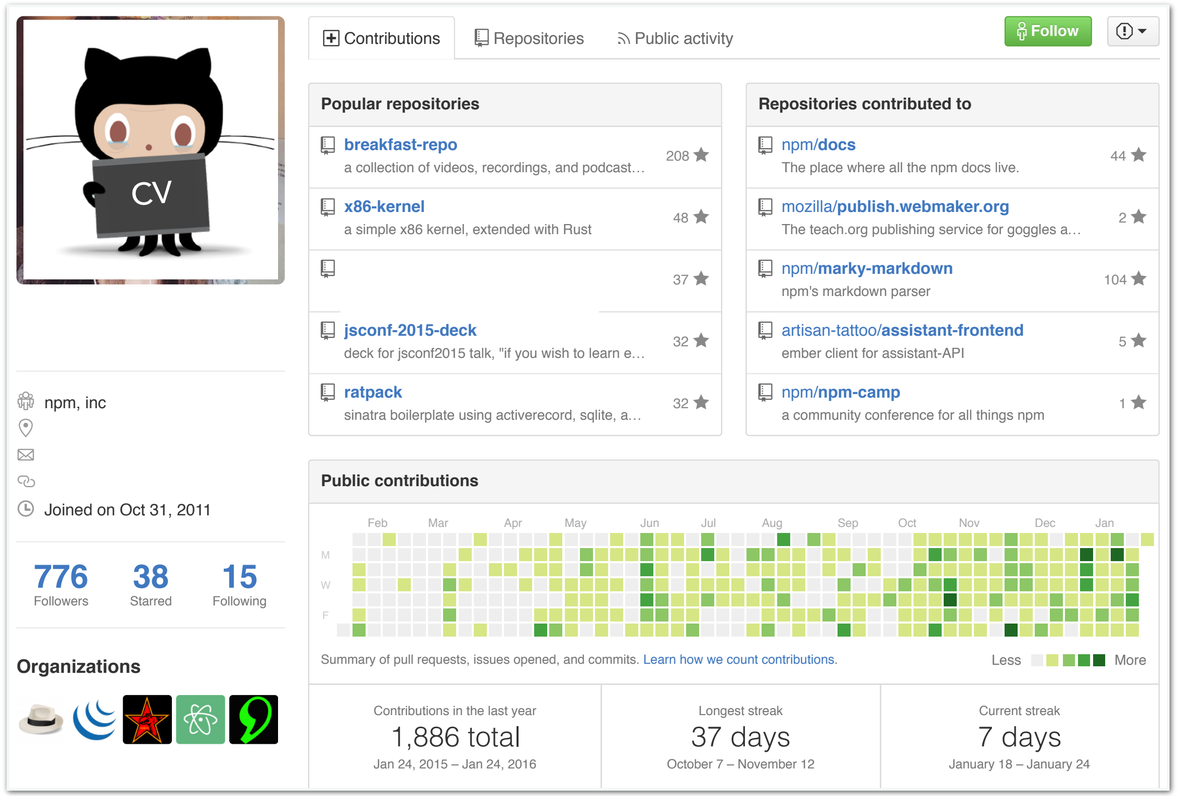
Find relevant mentorship

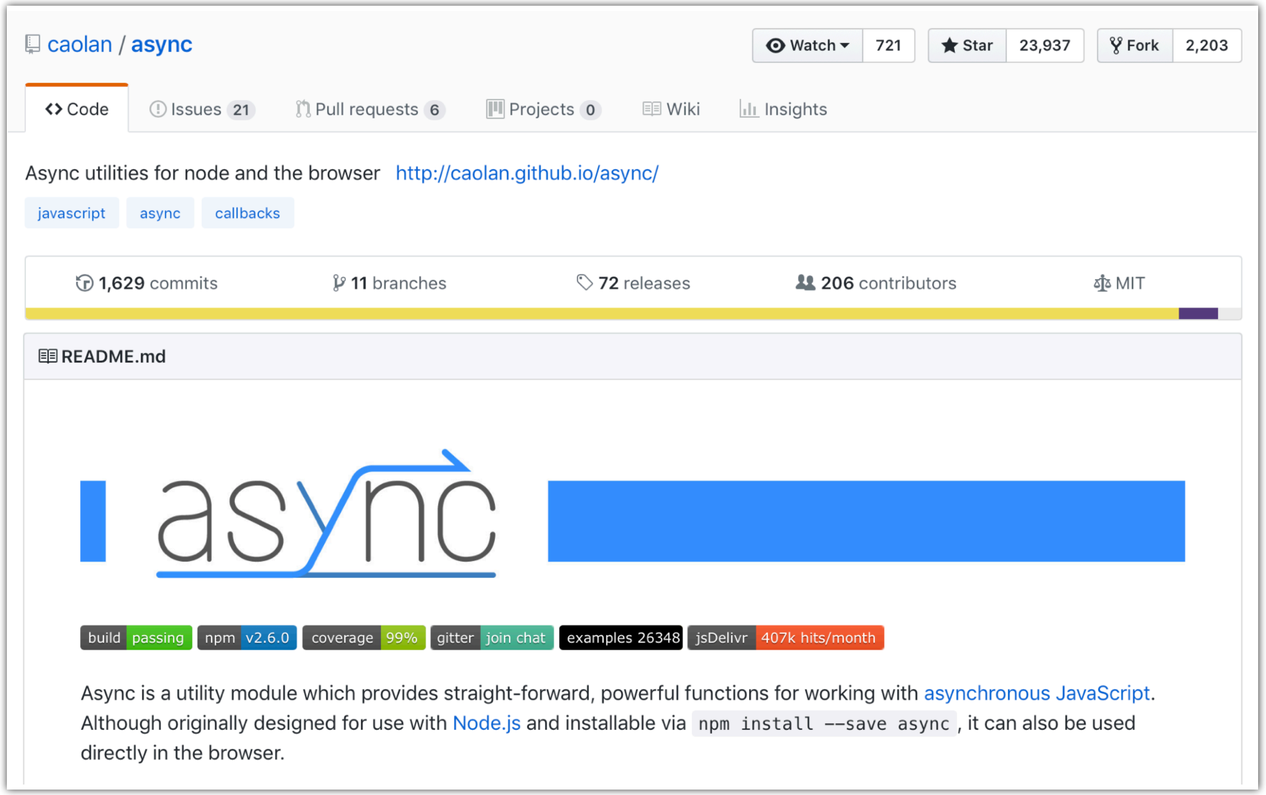


Signal social capital moderators

Leveraging signals

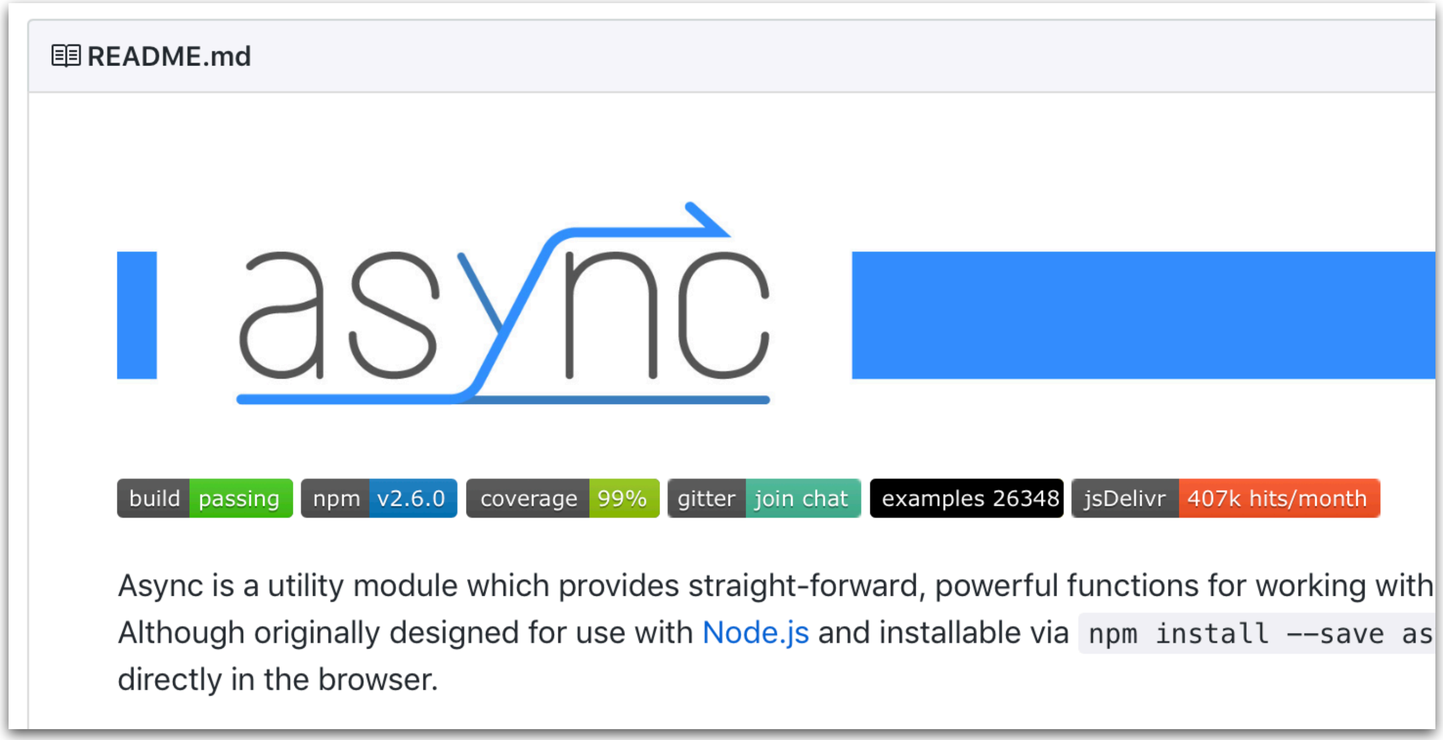
Transparency is already a defining characteristic of the environment





Signals are customizable

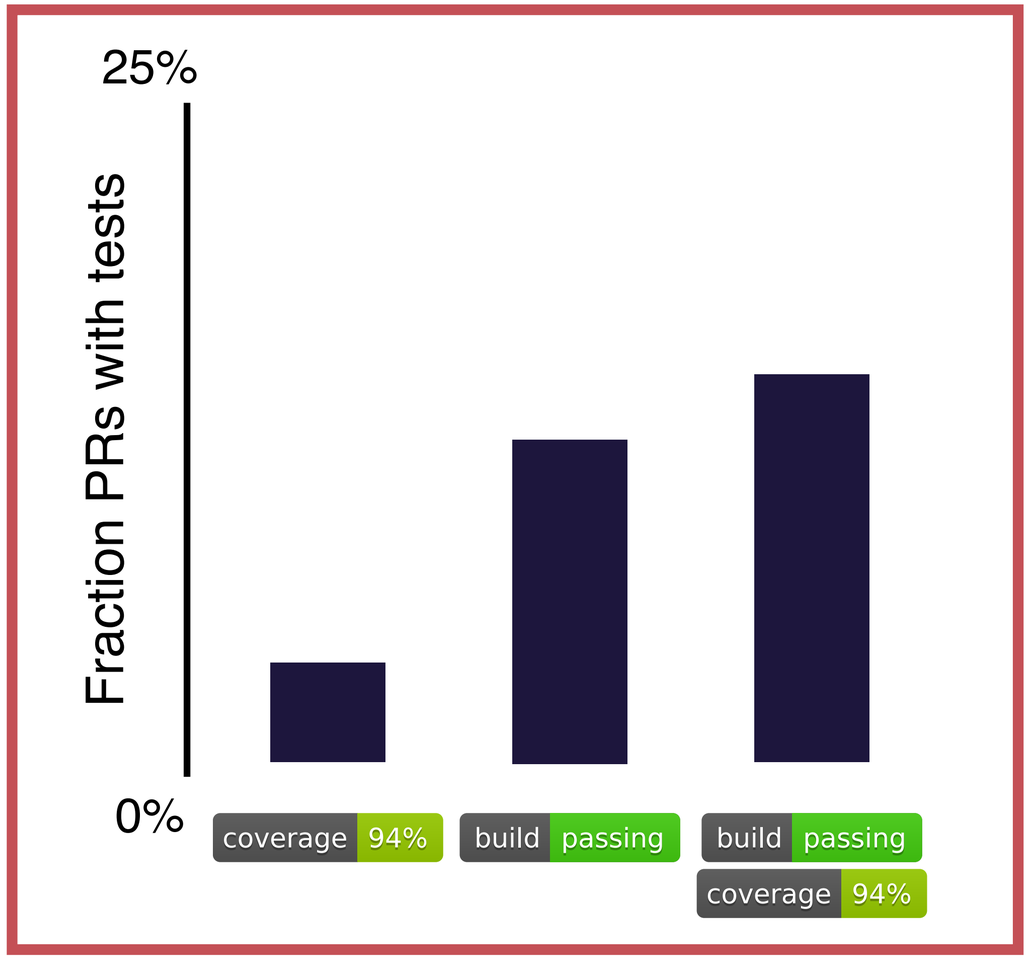
- E.g., repository badges

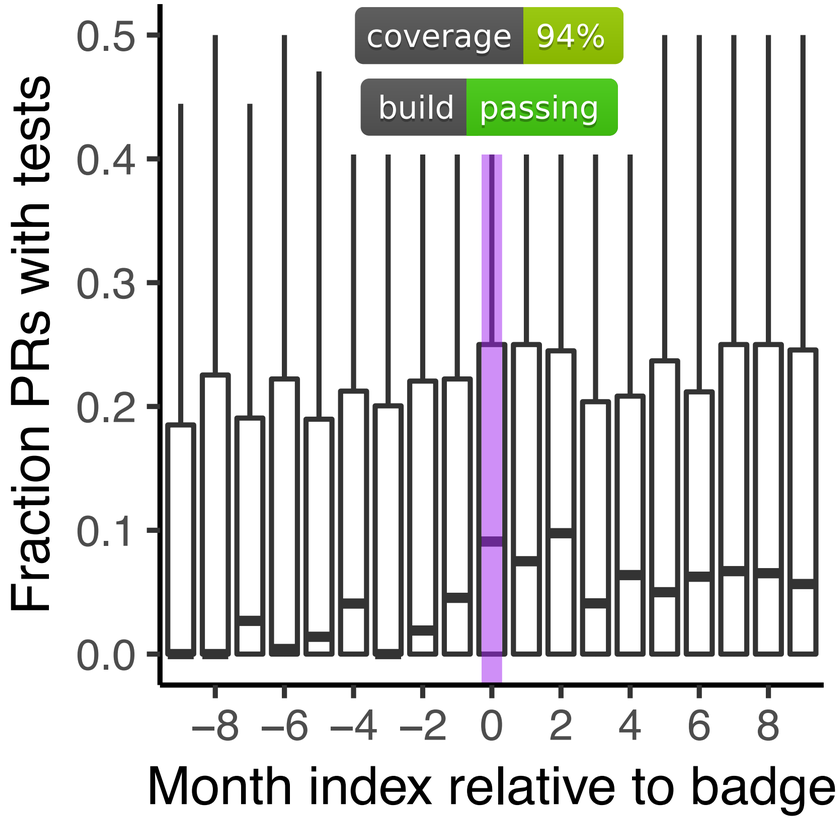


- Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the npm Ecosystem. Trockman, A., Zhou, S., Kästner, C., and Vasilescu, B. *ICSE 2018*

Signals are effective at steering behavior

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





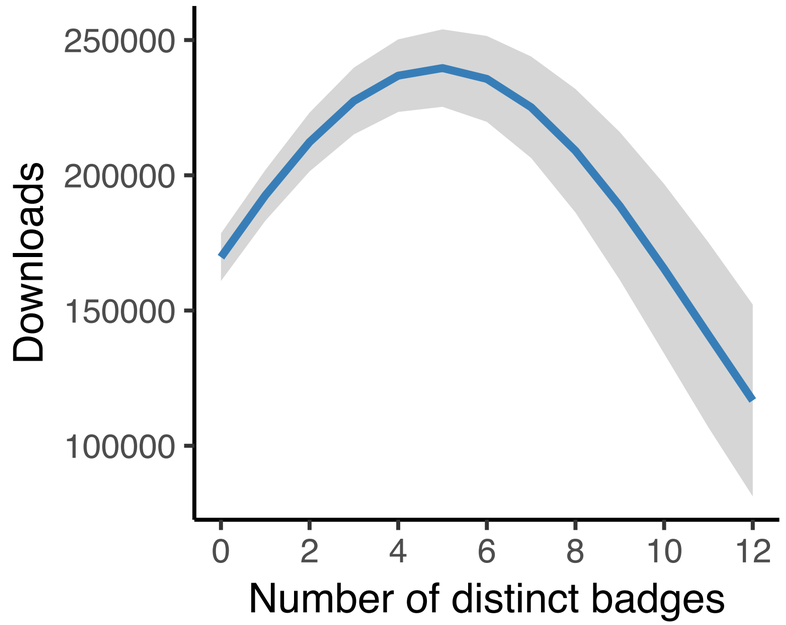
- Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the npm Ecosystem. Trockman, A., Zhou, S., Kästner, C., and Vasilescu, B. *ICSE 2018*

Hypothesis:
Signals can help the ecosystem self-regulate


!
CAUTION

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

new media & society
14(4) 669–683

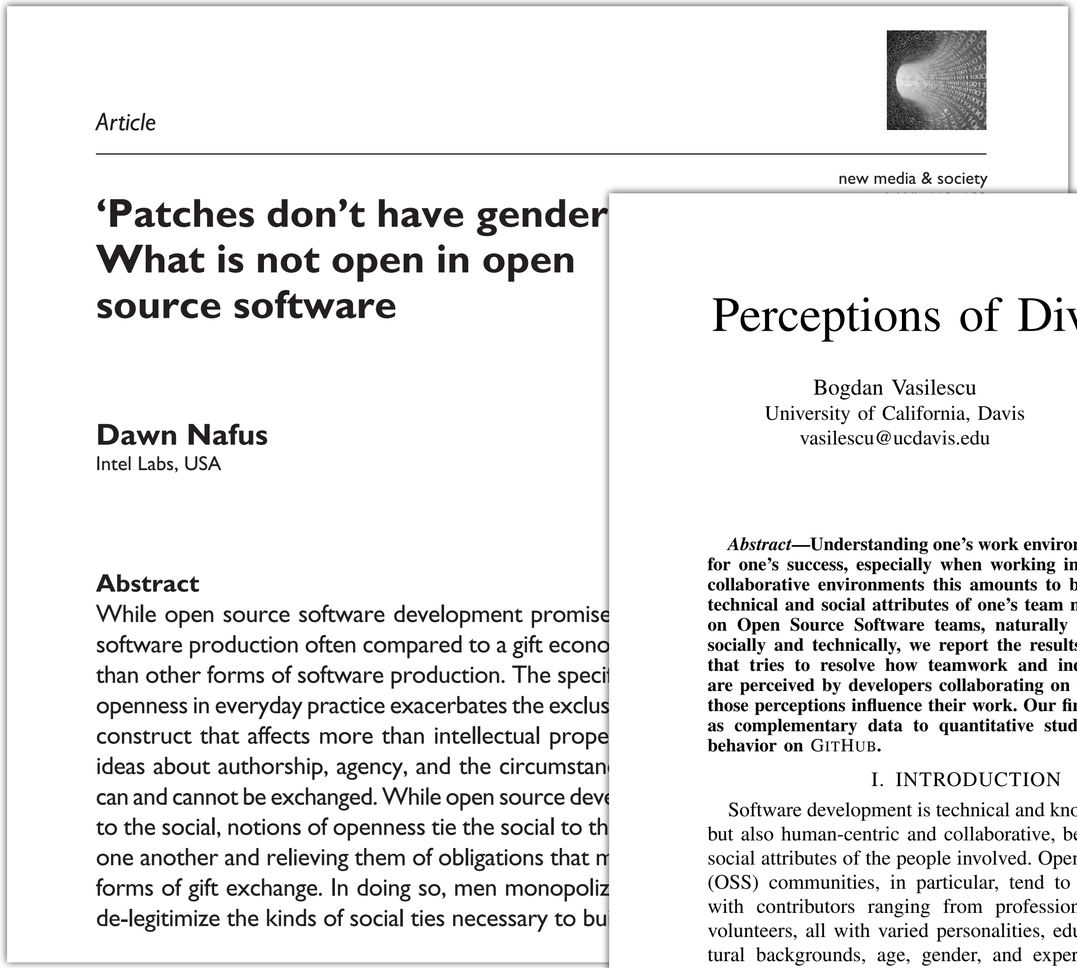
DOI: 10.1177/1461444811422887
nms.sagepub.com
SAGE

‘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 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”



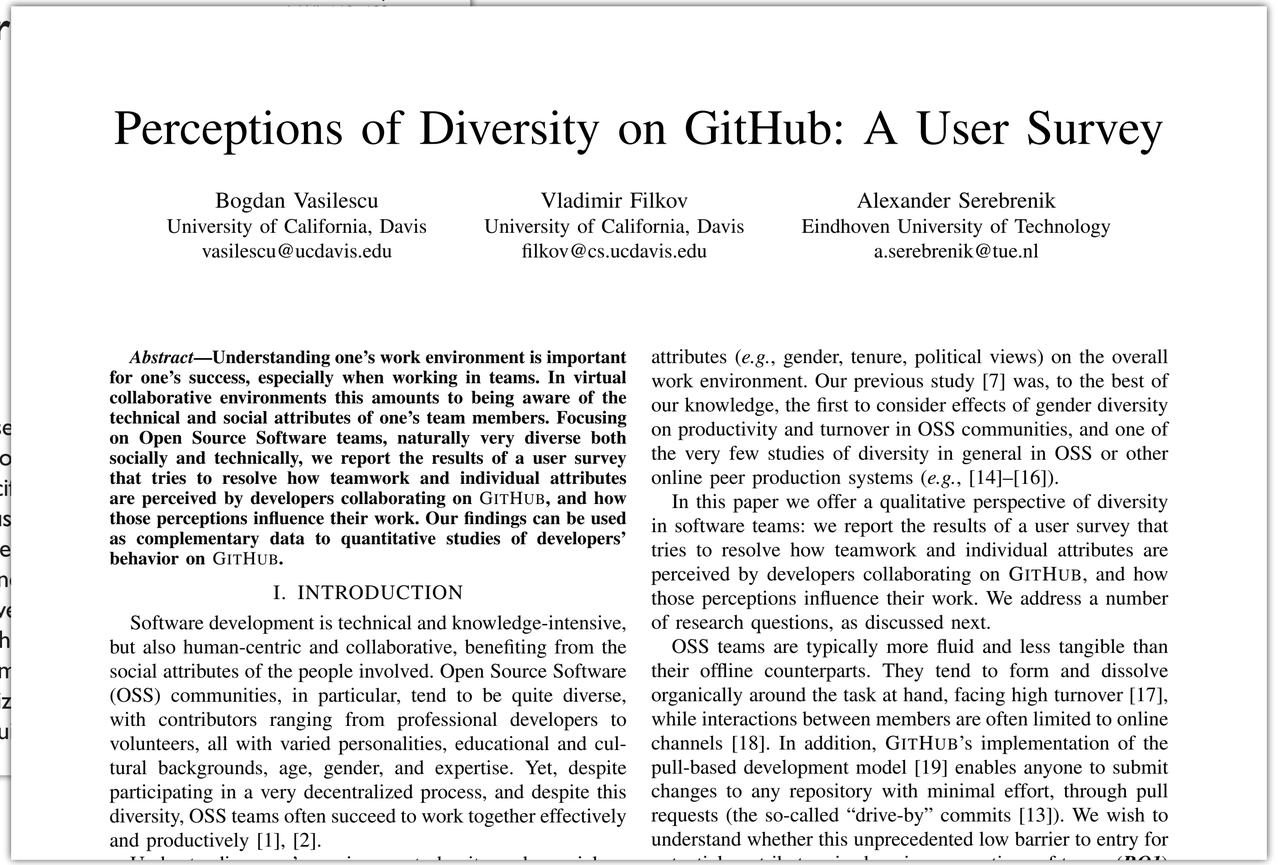
Article

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 promise
software production often compared to a gift econo
than other forms of software production. The speci
openness in everyday practice exacerbates the exclus
construct that affects more than intellectual prope
ideas about authorship, agency, and the circumstan
can and cannot be exchanged. While open source deve
to the social, notions of openness tie the social to th
one another and relieving them of obligations that m
forms of gift exchange. In doing so, men monopoliz
de-legitimize the kinds of social ties necessary to bu



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

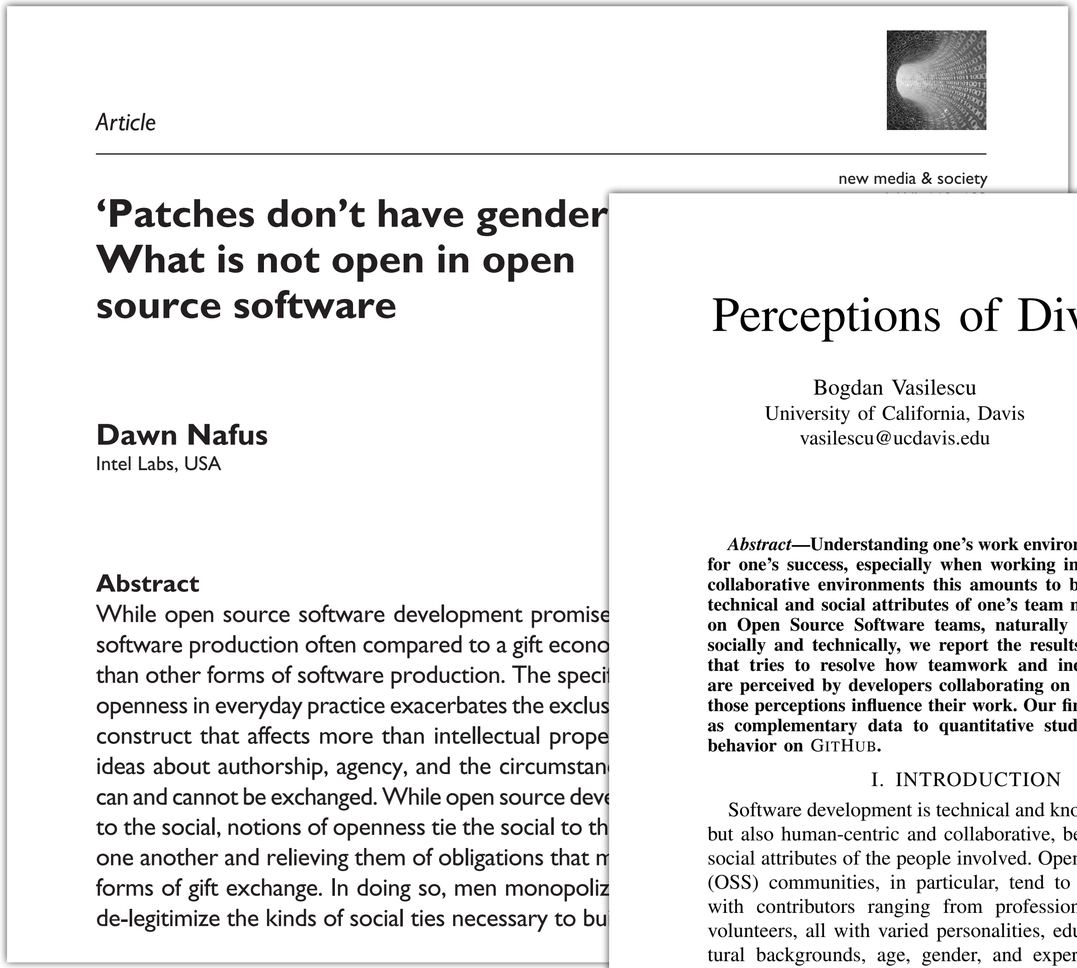
Software development is technical and knowledge-intensive, but also human-centric and collaborative, benefiting from the social attributes of the people involved. Open Source Software (OSS) communities, in particular, tend to be quite diverse, with contributors ranging from professional developers to 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



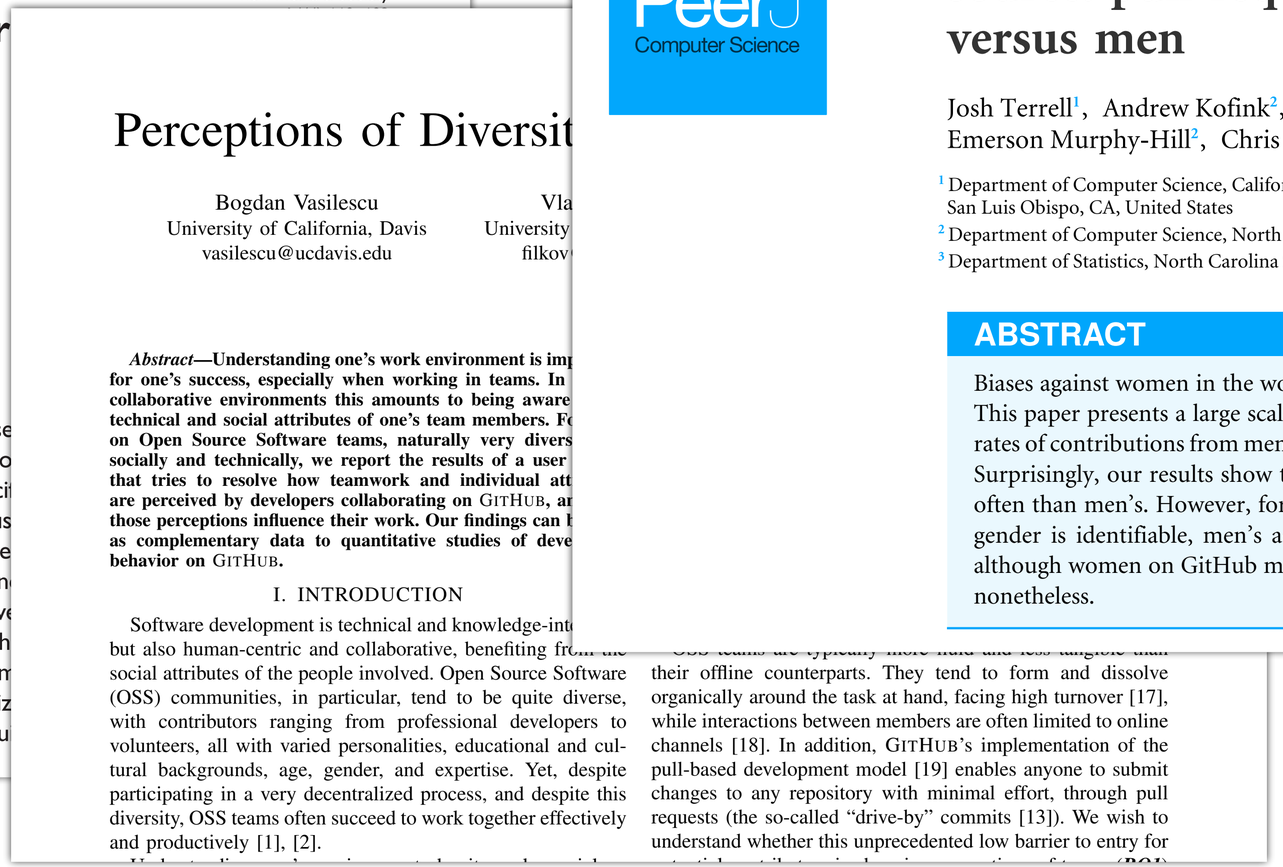
Article

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 promise
software production often compared to a gift econo
than other forms of software production. The speci
openness in everyday practice exacerbates the exclus
construct that affects more than intellectual prope
ideas about authorship, agency, and the circumstan
can and cannot be exchanged. While open source deve
to the social, notions of openness tie the social to th
one another and relieving them of obligations that m
forms of gift exchange. In doing so, men monopoliz
de-legitimize the kinds of social ties necessary to bu



Perceptions of Diversit

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

Vla
University
filkov

Abstract—**Understanding one’s work environment is imp
for one’s success, especially when working in teams. In
collaborative environments this amounts to being aware
technical and social attributes of one’s team members. Fo
on Open Source Software teams, naturally very divers
socially and technically, we report the results of a user
that tries to resolve how teamwork and individual att
are perceived by developers collaborating on GitHub, ar
those perceptions influence their work. Our findings can h
as complementary data to quantitative studies of deve
behavior on GitHub.**

I. INTRODUCTION

Software development is technical and knowledge-int
but also human-centric and collaborative, benefiting fro
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 cul-
tural 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].

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

PeerJ Computer Science

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

Josh Terrell[1], Andrew Kofink[2], Justin Middleton[2], Clarissa Rainear[2], Emerson Murphy-Hill[2], Chris Parnin[2] and Jon Stallings[3]

[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.

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*"

stackoverflow 5.8% | github SOCIAL CODING ~5%

open source 10.9% | Google 18% | 16.6%

In summary:
Many possible interventions

Missing: THEORY

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

Still,

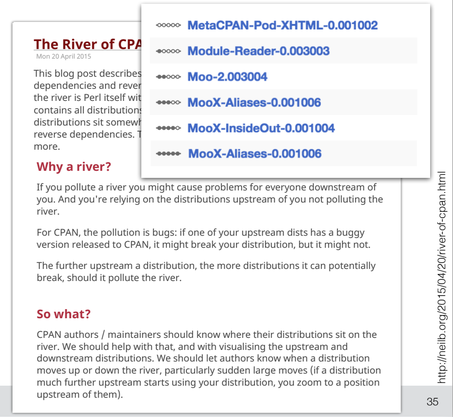
Signals could help

- disincentivize bad be
- match people to suit
- match people to suit
- …

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



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)

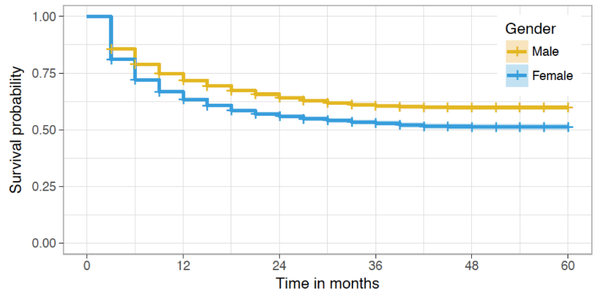
The tone of the community is an important factor in both interviews and model.

As
an
int

• 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*