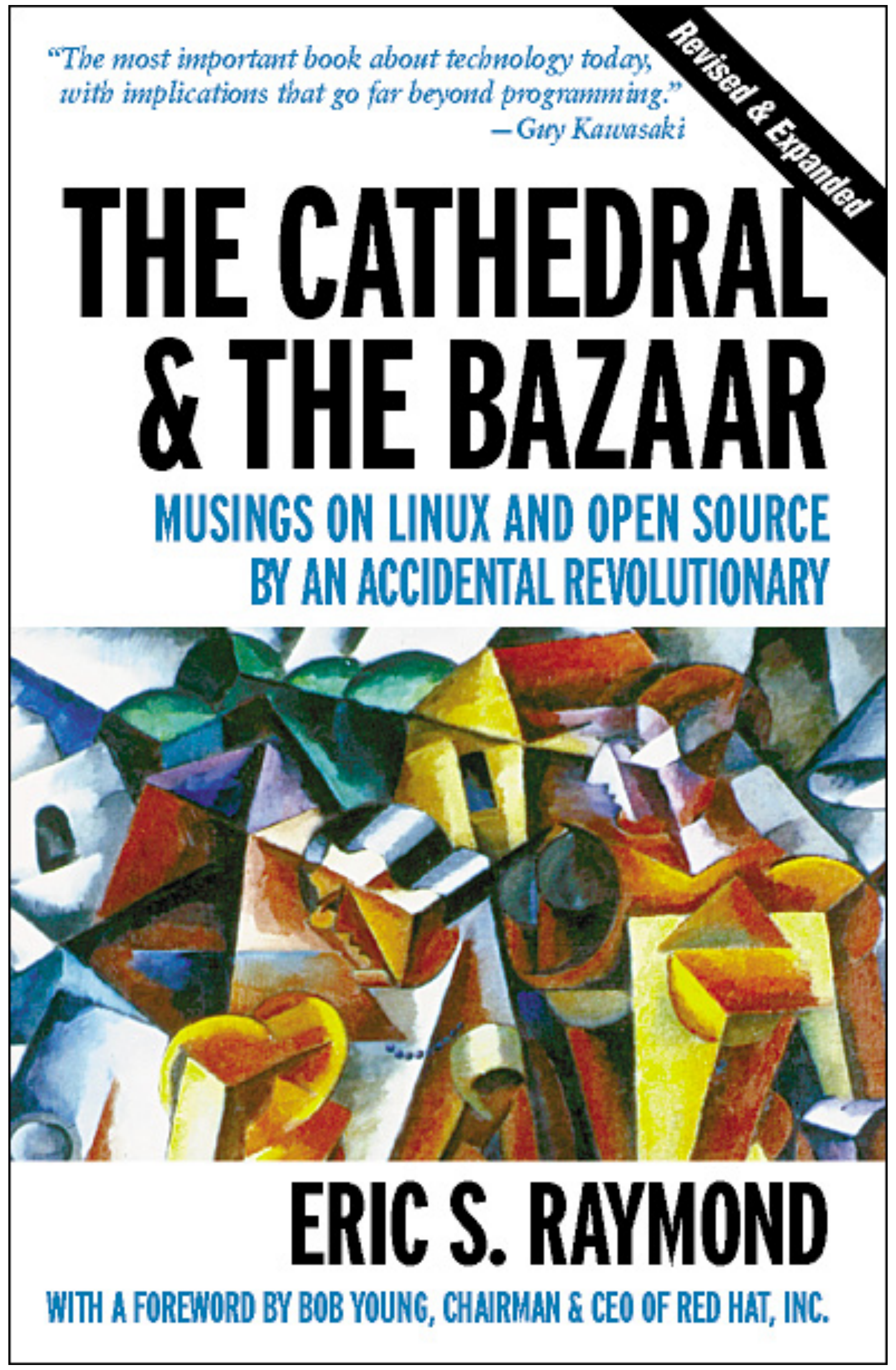
University of Zürich, March 14, 2019

Sustaining open source digital infrastructure

Bogdan Vasilescu
@b_vasilescu

Open source software: from curiosity to digital infrastructure

1999 → 2016



"The most important book about technology today, with implications that go far beyond programming." —Guy Kawasaki

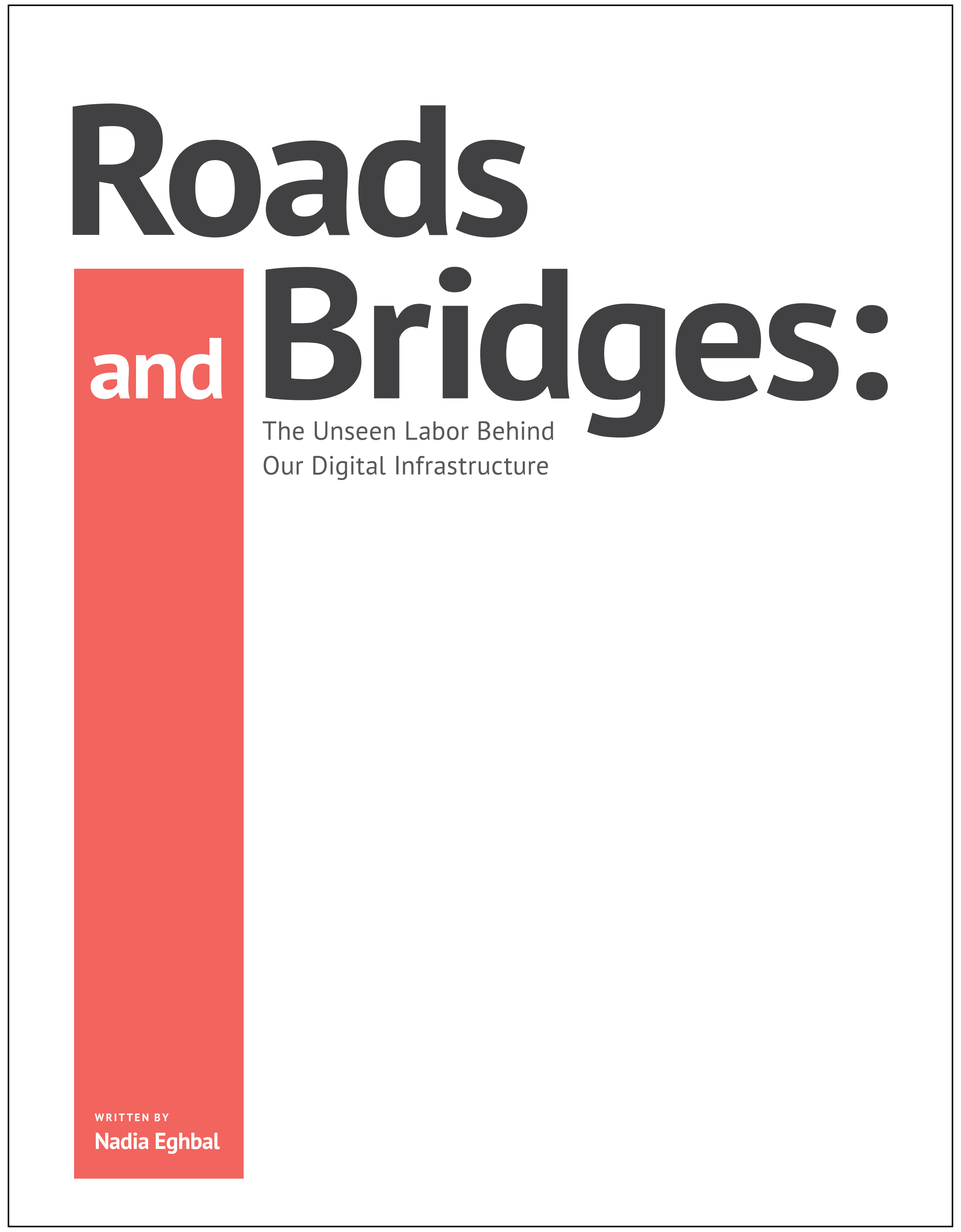
Revised & Expanded

THE CATHEDRAL & THE BAZAAR

MUSINGS ON LINUX AND OPEN SOURCE BY AN ACCIDENTAL REVOLUTIONARY

ERIC S. RAYMOND

WITH A FOREWORD BY BOB YOUNG, CHAIRMAN & CEO OF RED HAT, INC.



Roads and Bridges:

The Unseen Labor Behind Our Digital Infrastructure

WRITTEN BY Nadia Eghbal

- 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

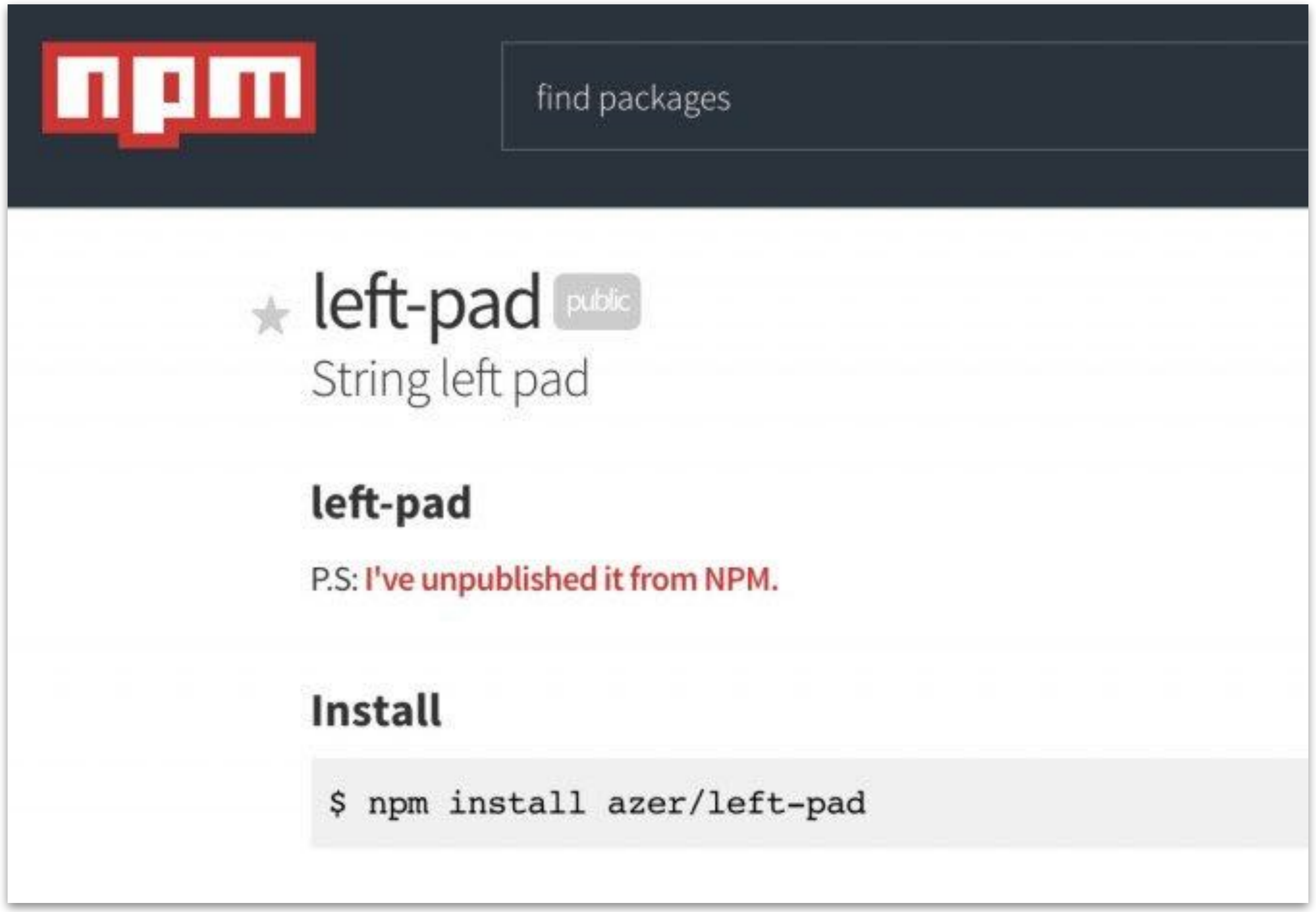
Economists: open source as “digital dark matter”

I.e., important but mostly invisible

- The installations of the Apache web server valued at $7 to $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



left-pad

String left pad

left-pad

P.S: I've unpublished it from NPM.

Install

```
$ npm install azer/left-pad
```

Open source needs a steady supply of time and effort by contributors

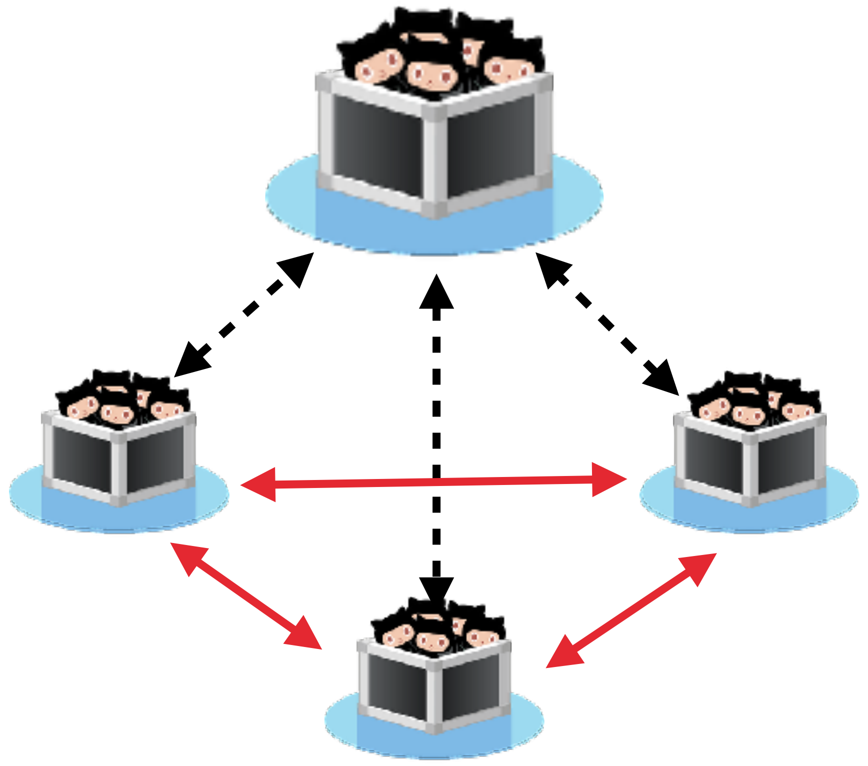
But that is harder today than ever before

… because of how open source has changed

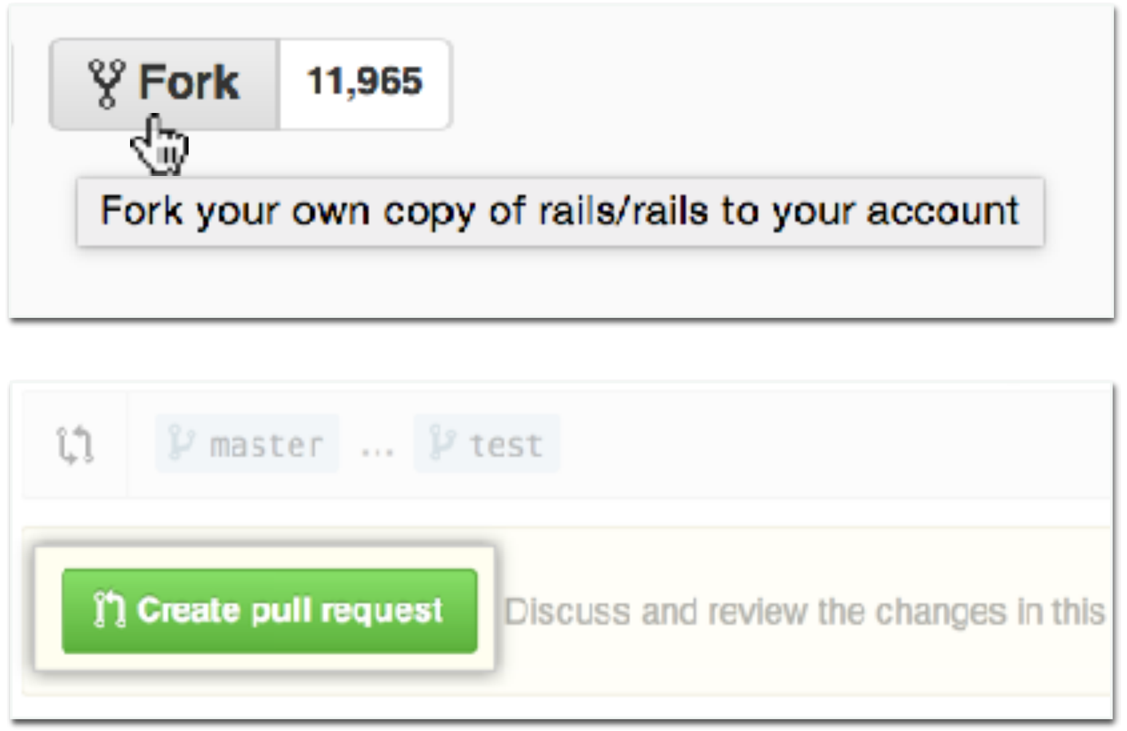
Today: more problems than solutions

Change:
GitHub as a standardized place to collaborate on code

- Git version control



- GitHub UI



Fork 11,965

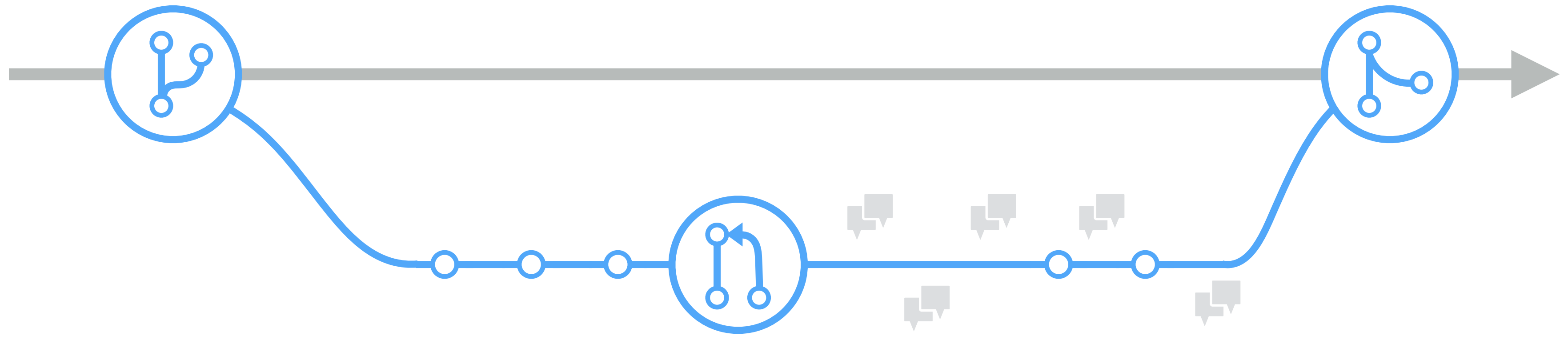
Fork your own copy of rails/rails to your account

master ... test

Create pull request

Discuss and review the changes in this

- 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

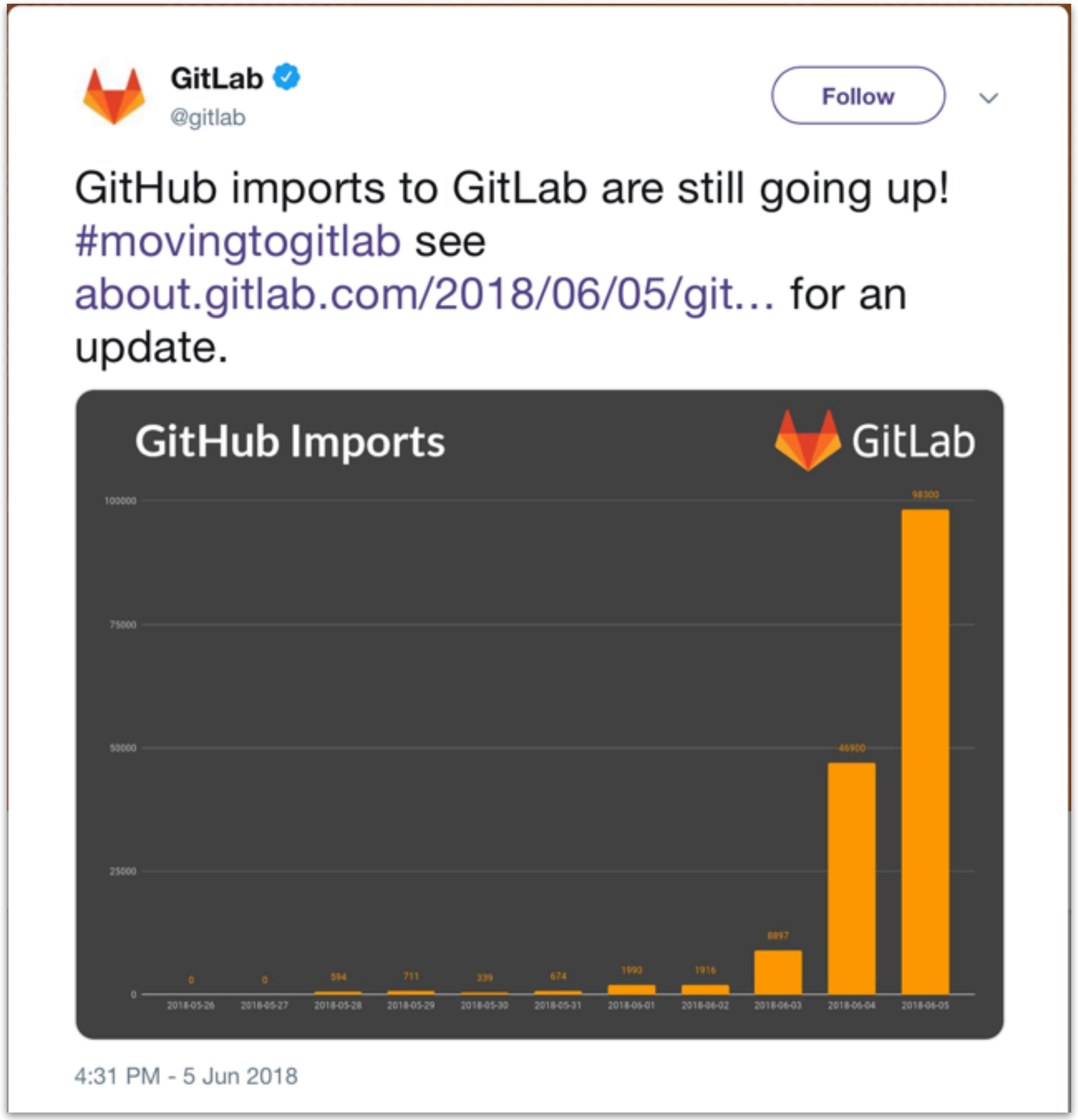
31 million users

(November 2018)

Bitbucket

6 million users

(March 2019)



GitLab @gitlab

Follow

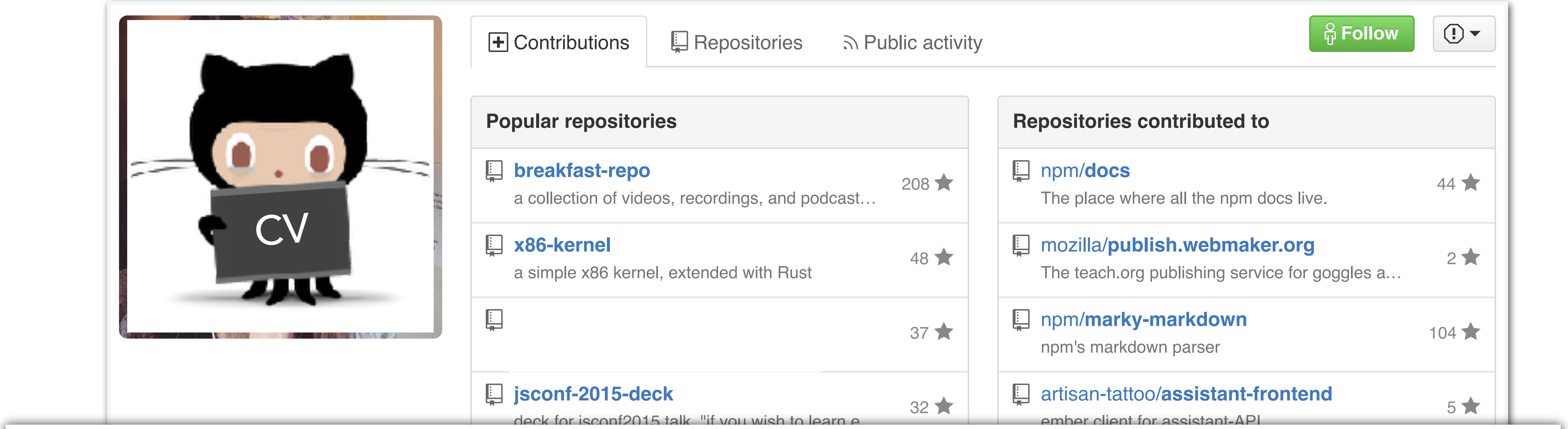
GitHub imports to GitLab are still going up! #movingtogitlab see about.gitlab.com/2018/06/05/git… for an update.

4:31 PM - 5 Jun 2018

Change: High level of transparency

- Clear awareness of the audience, which influences how people behave
  - GitHub is like being onstage
    - (Dabbish et al. 2012)

- Signaling mechanisms
  - Individual expertise, to potential employers
    - (Marlow et al. 2013), (Marlow and Dabbish 2013)
  - Project qualities, to contributors and users
    - (Trockman et al. 2018)

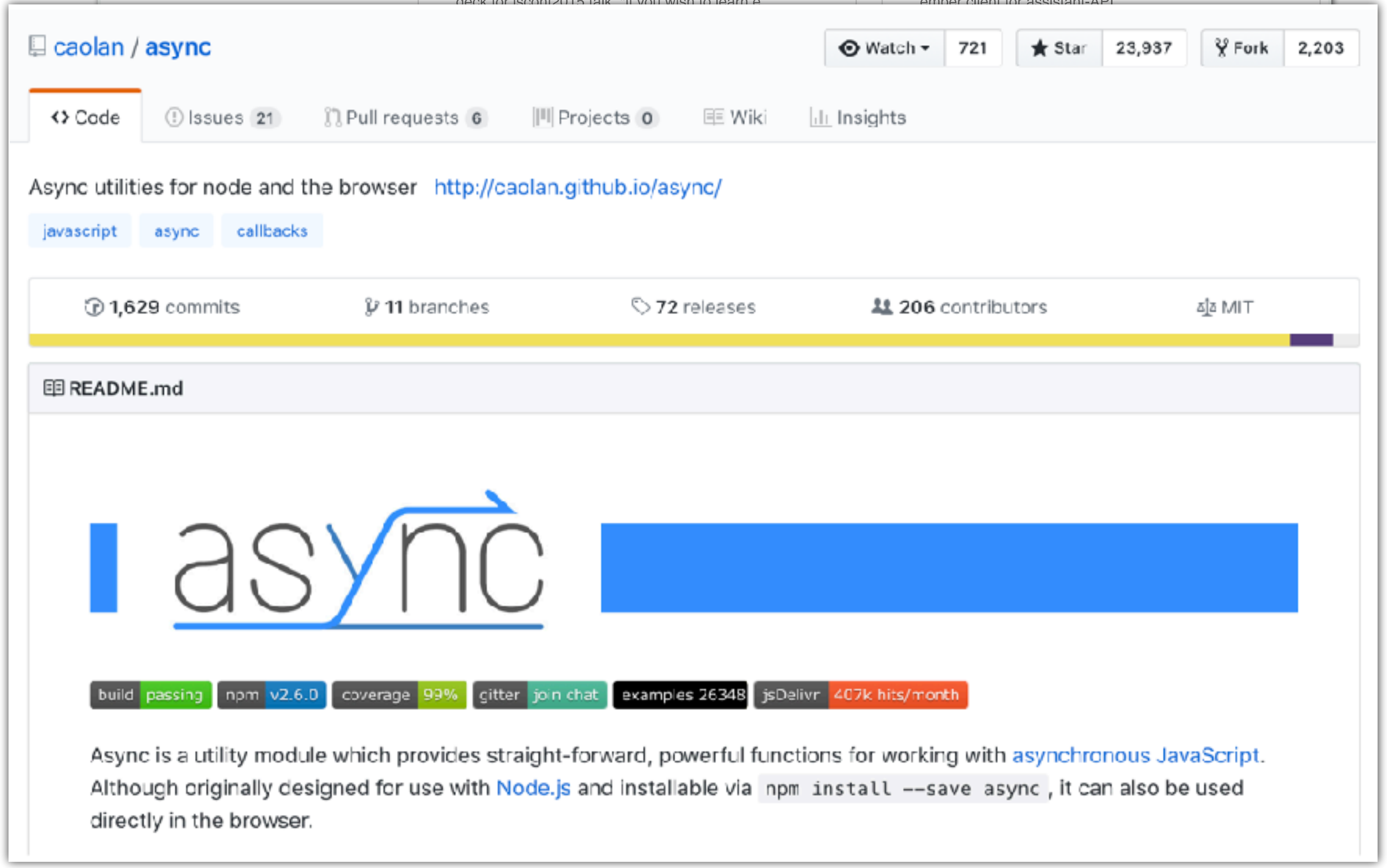


Popular repositories

- breakfast-repo — a collection of videos, recordings, and podcast… — 208 ★
- x86-kernel — a simple x86 kernel, extended with Rust — 48 ★
- 37 ★
- jsconf-2015-deck — deck for jsconf2015 talk. "if you wish to learn e… — 32 ★

Repositories contributed to

- npm/docs — The place where all the npm docs live. — 44 ★
- mozilla/publish.webmaker.org — The teach.org publishing service for goggles a… — 2 ★
- npm/marky-markdown — npm's markdown parser — 104 ★
- artisan-tattoo/assistant-frontend — ember client for assistant-API — 5 ★



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

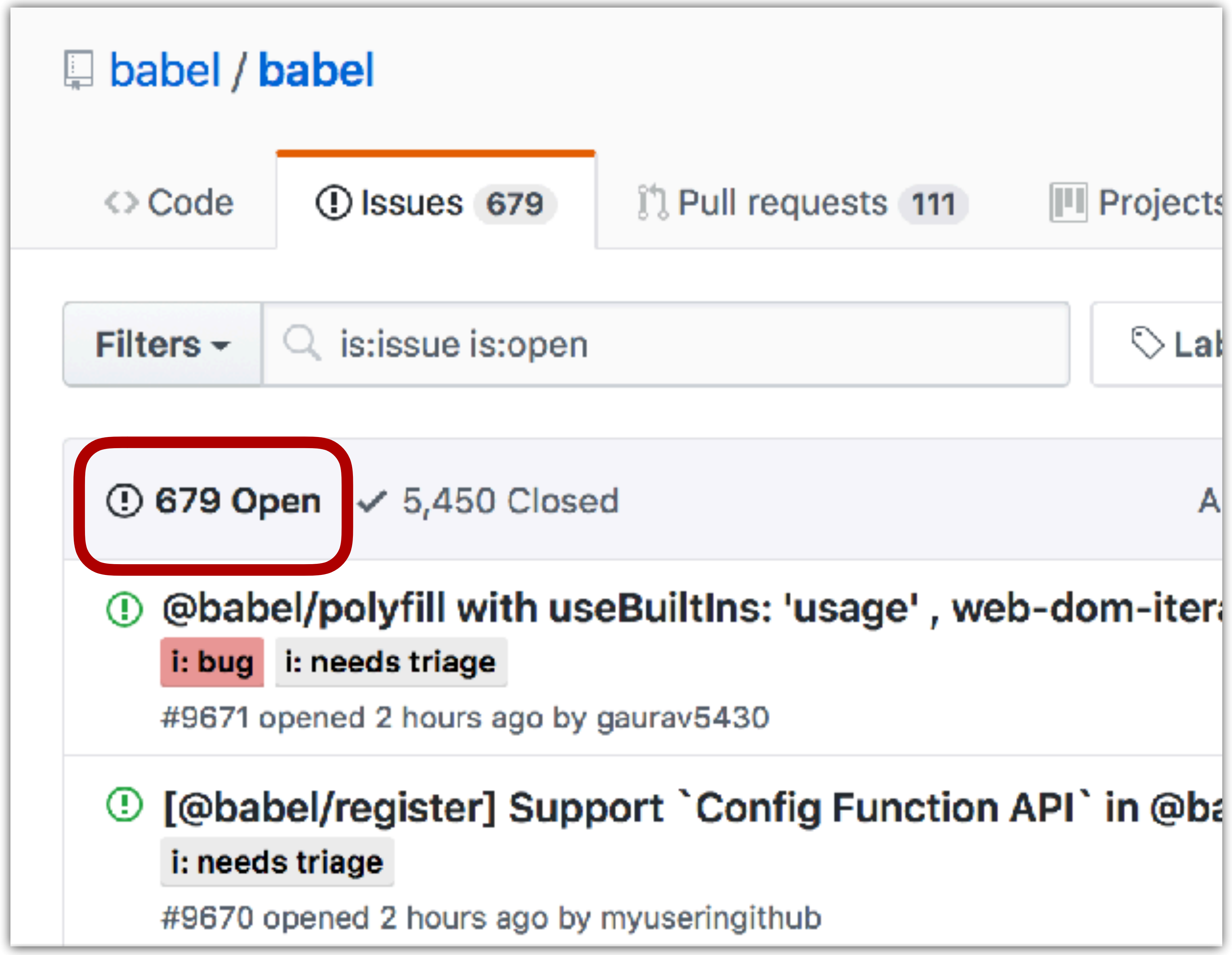
1,629 commits | 11 branches | 72 releases | 206 contributors | MIT

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.

- 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*

Challenge: 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.”*



babel / babel

is:issue is:open

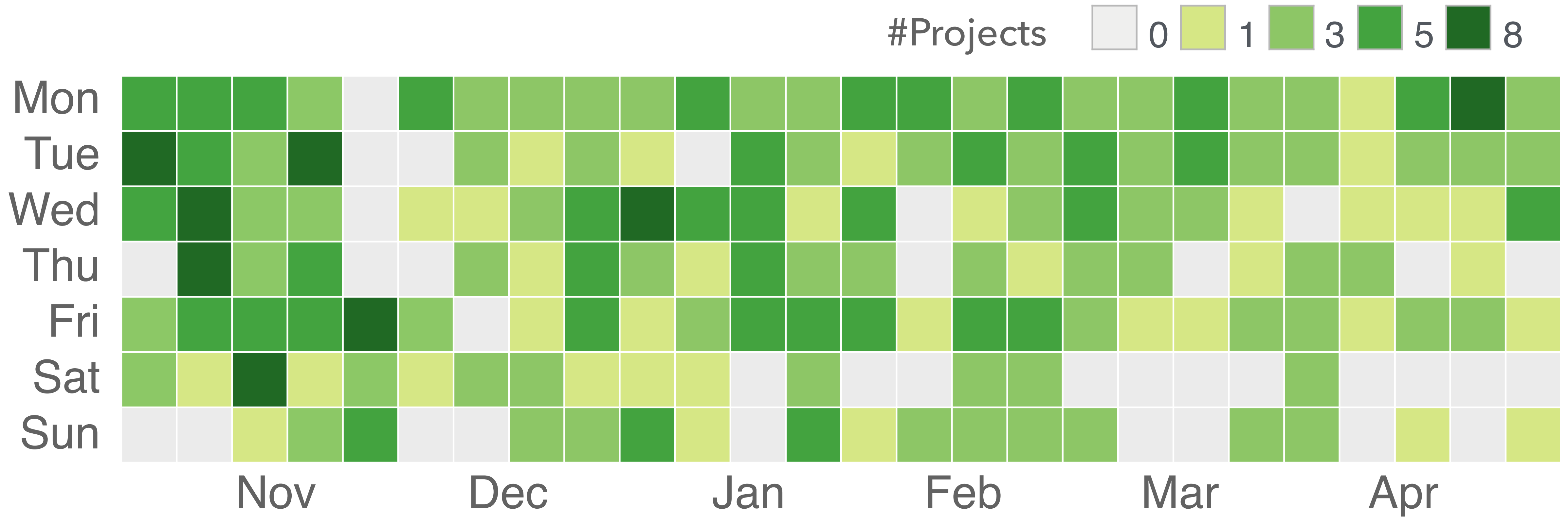
679 Open ✓ 5,450 Closed

@babel/polyfill with useBuiltIns: 'usage' , web-dom-iter
i: bug i: needs triage
#9671 opened 2 hours ago by gaurav5430

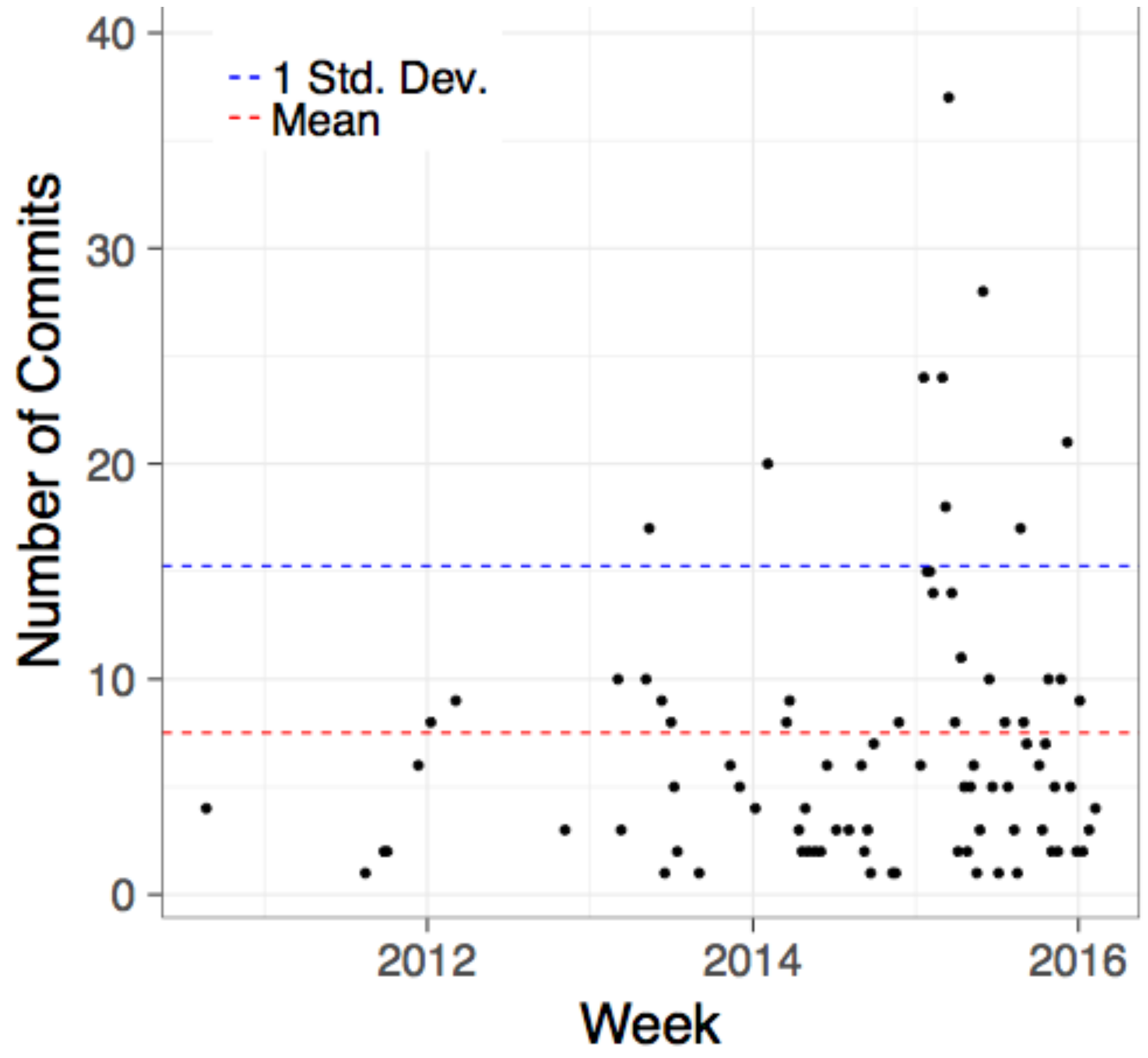
[@babel/register] Support `Config Function API` in @ba
i: needs triage
#9670 opened 2 hours ago by myuseringithub

Challenge:
High-workload, potentially high-stress environment

- Working on many projects concurrently
  - (25 Nov 2013 — 18 May 2014)



- Periods with significantly 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*

Challenge: Low demographic diversity

- 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*"

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

- Gender representation reality

stackoverflow 5.8%

github SOCIAL CODING ~5%

open source 10.9%

Google 18%

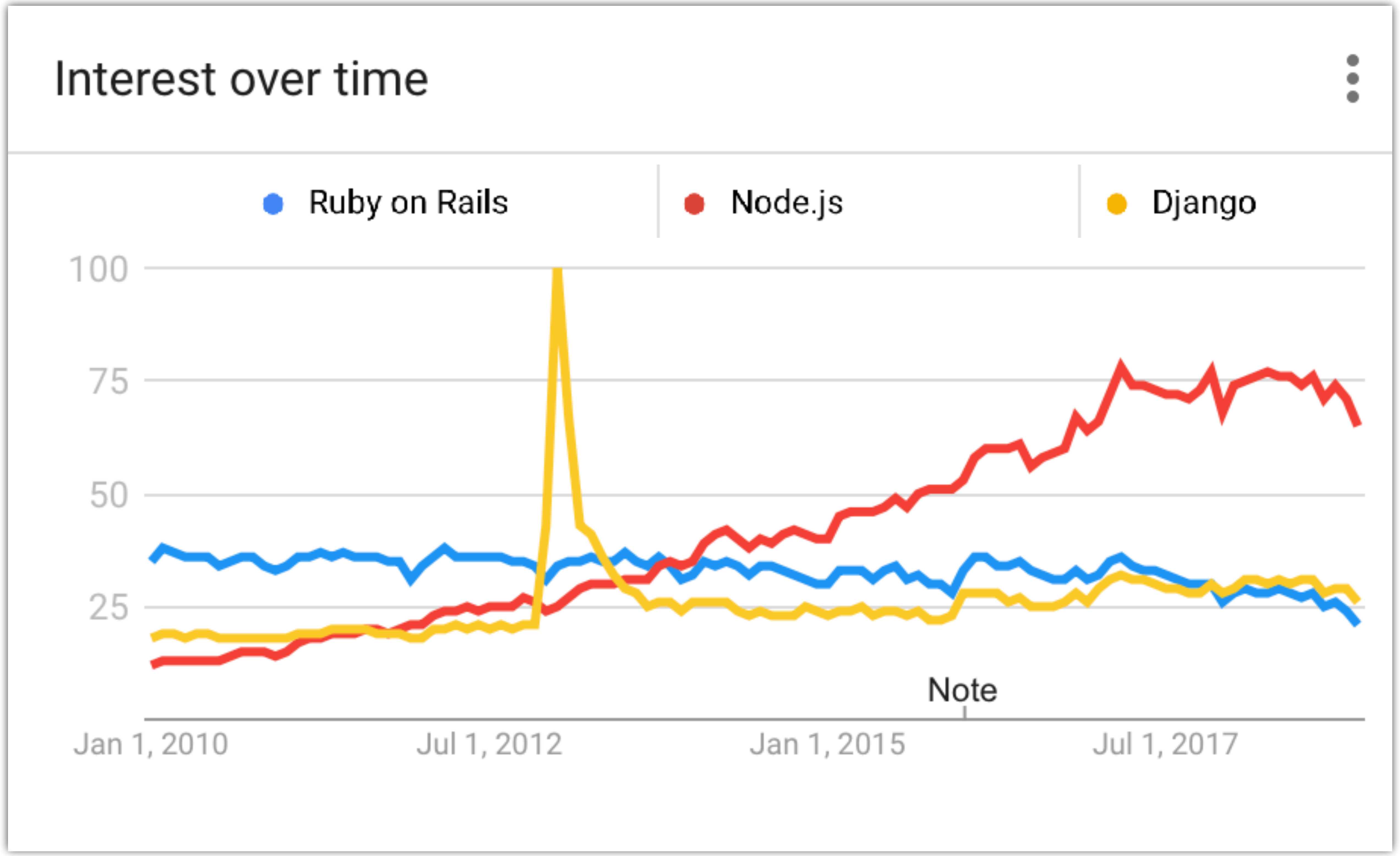
16.6%

- 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

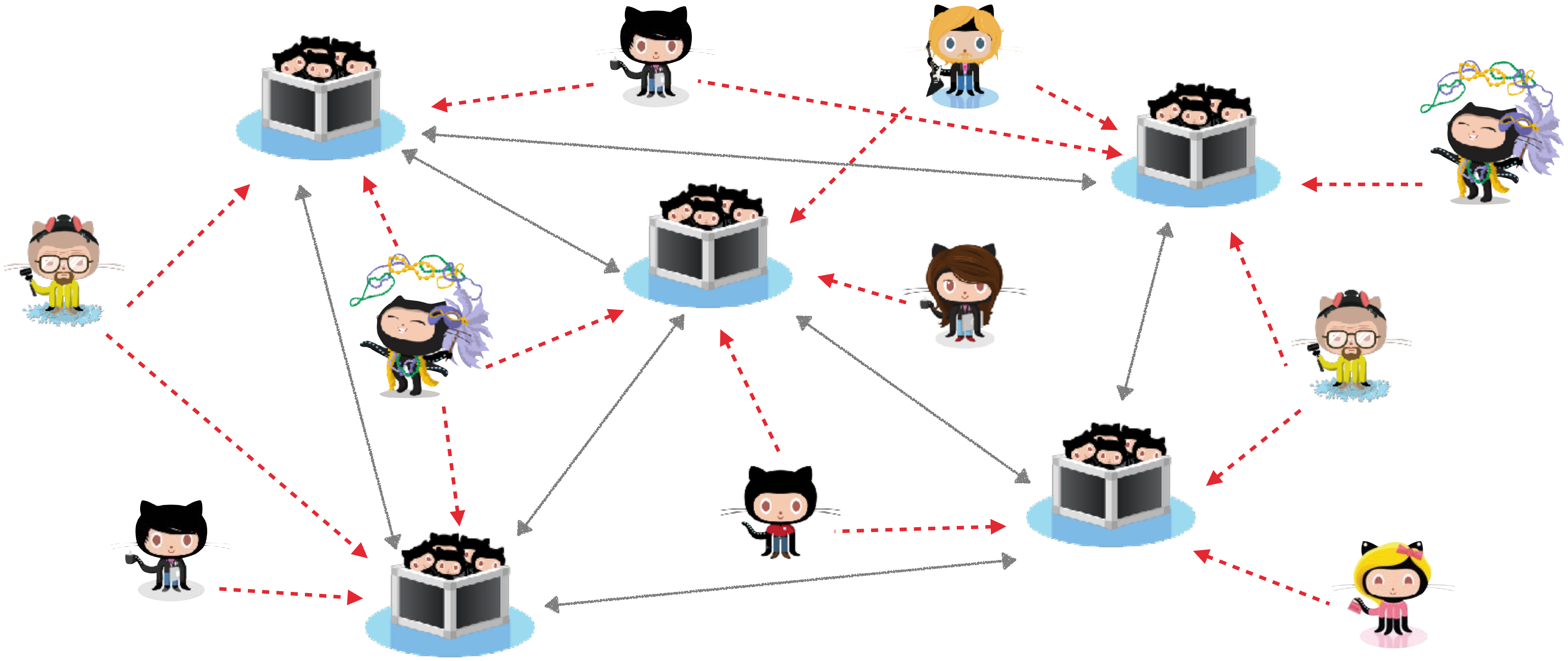
Challenge: Rapid evolution

- Hard to attract and retain contributors unless project is new and exciting
  - Interviewee looking at GitHub stars [ongoing research]:
  - *"It doesn't look like it's popular enough to really have enough impact to warrant your time"*



Google Trends

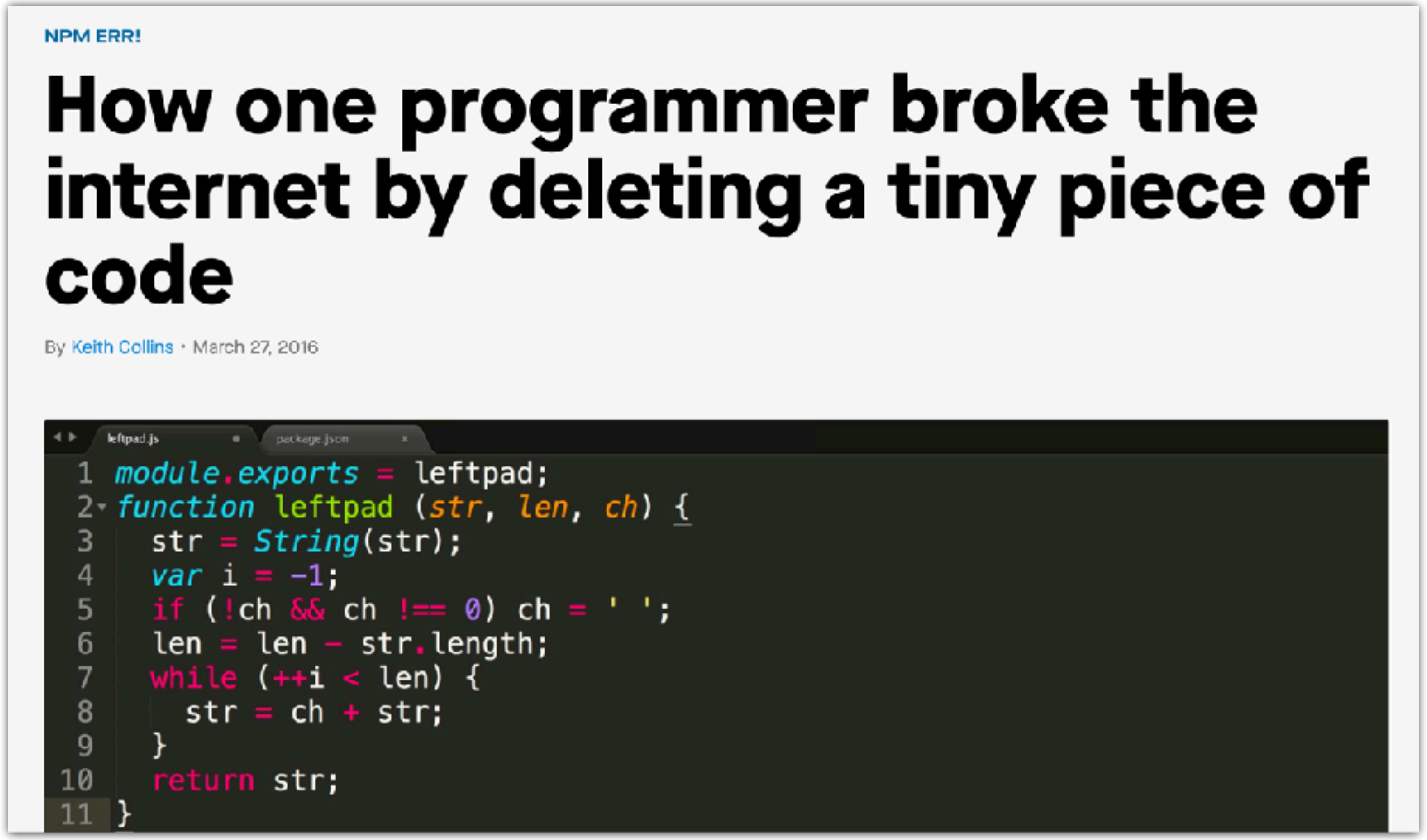
Change: Complex ecosystems of interdependencies



- Socio-technical environment: heterogeneous links

Challenge: 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

```
module.exports = leftpad;
function leftpad (str, len, ch) {
  str = String(str);
  var i = -1;
  if (!ch && ch !== 0) ch = ' ';
  len = len - str.length;
  while (++i < len) {
    str = ch + str;
  }
  return str;
}
```

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: Increasing commercialization and professionalization

- Historically
  - Community-based projects (Python, RubyGems, Twisted)

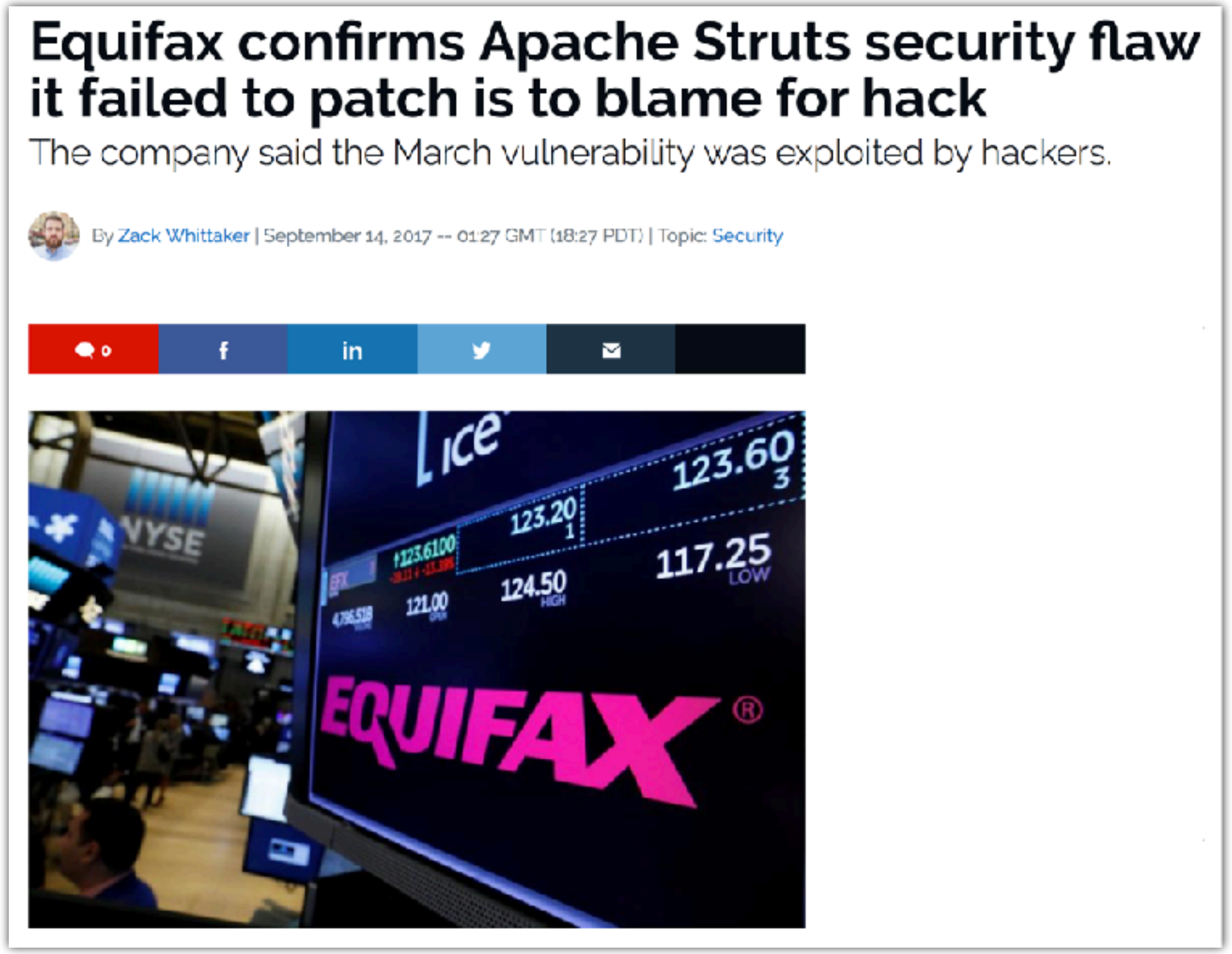
- Currently
  - Lots of commercial involvement
    - Companies (Go - Google, React - Facebook, Swift - Apple)
    - Startups (Docker, npm, Meteor)

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

http://opensourcesurvey.org/2017/

Challenge: 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

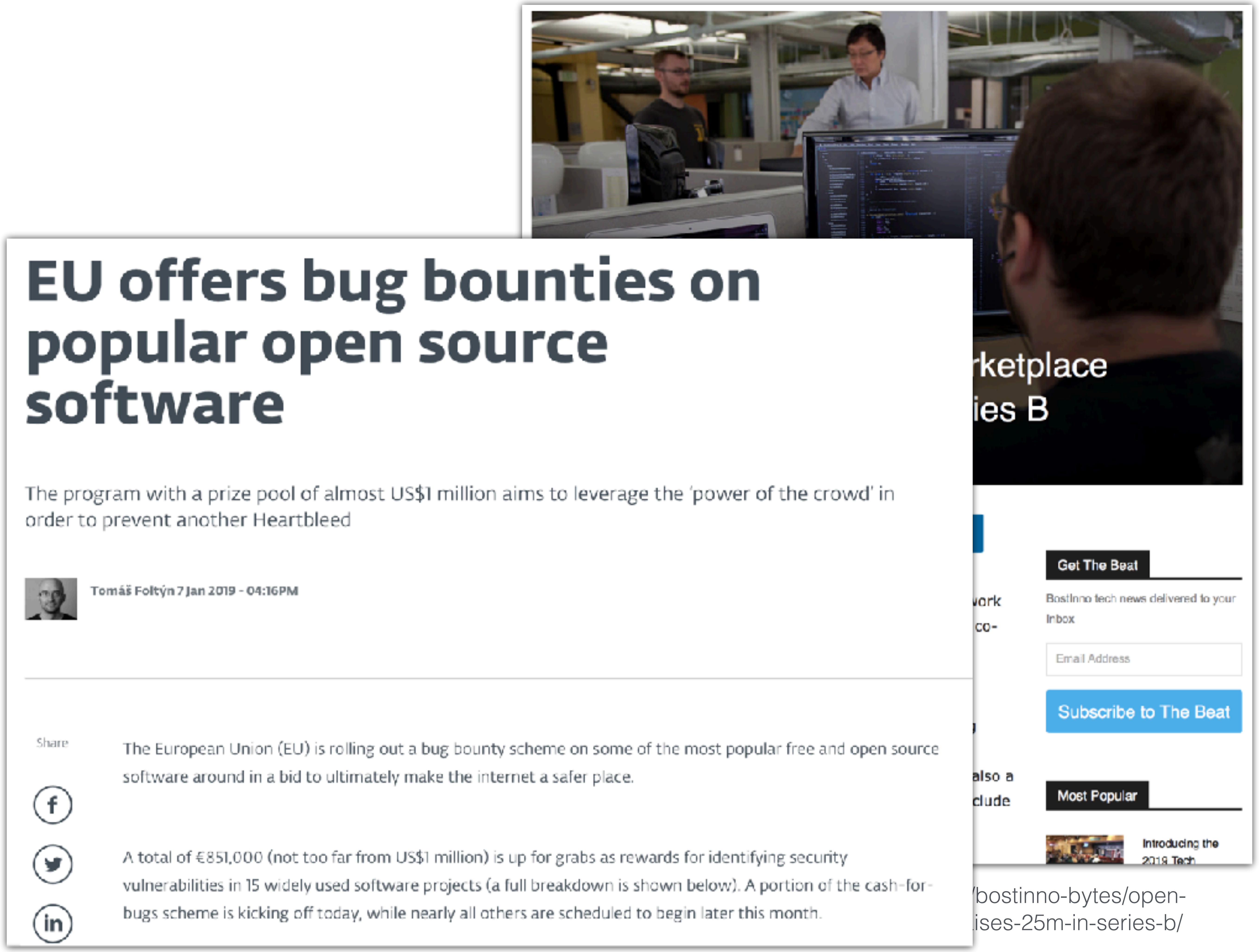
The company said the March vulnerability was exploited by hackers.

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/

Challenge: Money believed to have a corrupting influence

- Demotivating for contributors?
- Open source as public good:
  - Sponsoring development work may also benefit one's competitor, who may have not contributed anything



EU offers bug bounties on popular open source software

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

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.

/bostinno-bytes/open-
ises-25m-in-series-b/

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

Open source needs a steady supply of time and effort by contributors

But that is harder today than ever before

… because of how open source has changed

What can we do?

Two things are obvious (to me)

1. No individual person, company, or organization can address these problems alone
2. We need more science to understand:
    - which open source projects form digital infrastructure
    - how open source digital infrastructure is being used
    - how much and what kind of effort does each project need
    - how do project interdependencies impact sustainability
    - how do people choose which projects to contribute to
    - how to attract a more diverse pool of contributors
    - why do open source contributors disengage / how to retain them
    - which project-level practices and policies encourage contributions
    - how effective are the different support models / what are their side effects
    - how much can transparency help the ecosystem to self regulate

Great potential for quantitative empirical research: Big data in open source

1. FALSE POSITIVES
2. FALSE NEGATIVES
3. CONFOUNDS

	Reject Null Hyp.	Accept Null Hyp.
Null Hyp. TRUE	1	
Null Hyp. FALSE		2

HUGE SAMPLE SIZES:

- More stringent a priori about significance level
  → reduce False Positives
- Detect even small effects
  → reduce False Negatives
- Handle more degrees of freedom
  → control for Confounds

SEPARATE SIGNAL FROM NOISE:

- Quantify effect size
- Mix research methods
  - Theory: social sciences
  - Qualitative: case studies, user surveys, interviews, …
  - Quantitative: stats, data mining, …

VALIDATE DATA & MEASURES FIRST!

- Spot-checking

What can we do?

Two things are obvious (to me)

1. No individual person, company, or organization can address these problems alone
2. We need more science to understand:
   - which open source projects form digital infrastructure
   - how open source digital infrastructure is being used
   - how much and what kind of effort does each project need
   - **how do project interdependencies impact sustainability**
   - how do people choose which projects to contribute to
   - how to attract a more diverse pool of contributors
   - **why do open source contributors disengage** / how to retain them
   - which project-level practices and policies encourage contributions
   - how effective are the different support models / what are their side effects
   - how much can transparency help the ecosystem to self regulate

How do project interdependencies impact sustainability

[Valiev et al. ESEC/FSE 2018]

Leftpad 2.0: premises

- There is a Python package
  - only one non-trivial contributor
  - a few dozen commits in total
  - last commit over 5 months ago

  - ~15% of all packages depend on it
  - … including pip (package installer)

- Many factors external to a given project can impact its sustainability
  - upstream dependencies
  - funding agencies
  - external support
  - downstream communities
  - …
- It takes only one to break a project

Spoiler: External factors play an important role in the sustainability of open source projects

Methodology: mixed-methods empirical study

Data: Artifacts Evaluated acm Reusable Artifacts Evaluated acm Functional

70K PyPI packages

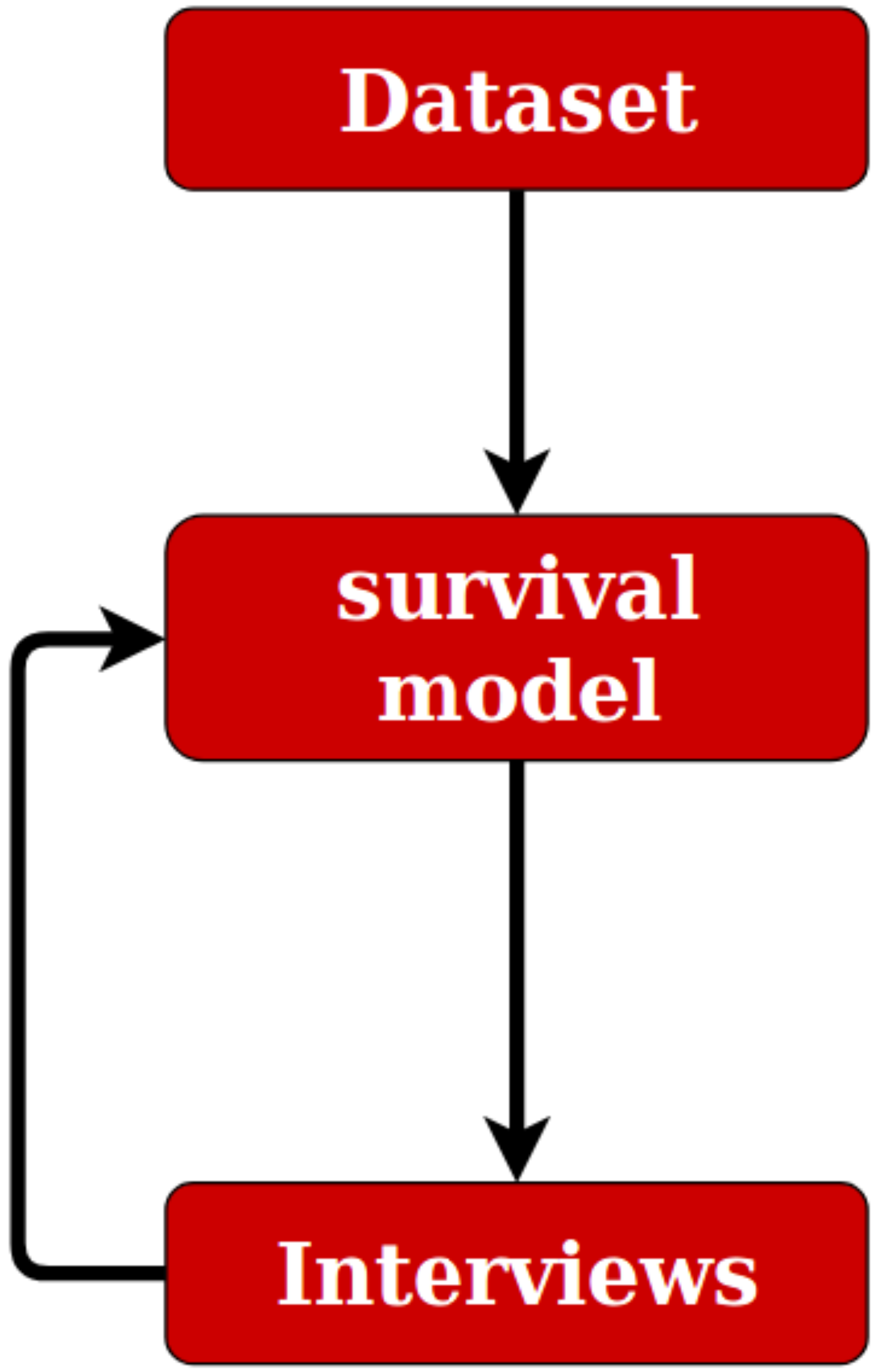
https://zenodo.org/record/1297925

Model:

Cox survival regression

Interviews:

10 project maintainers



Dataset

survival model

Interviews

Methodology: mixed-methods empirical study

Data:

70K PyPI packages

https://zenodo.org/record/1297925

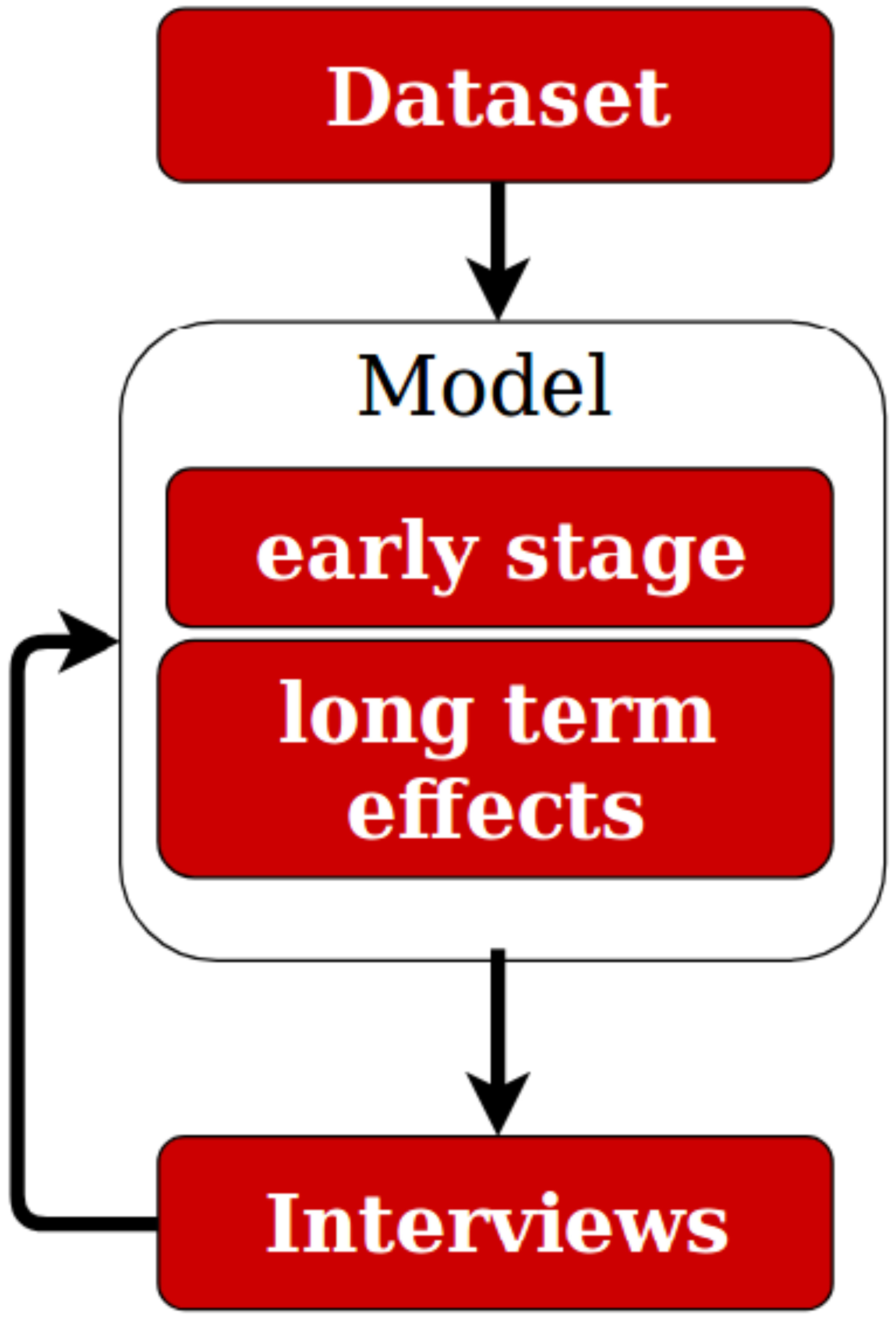
2-stage model:

Logistic Regression

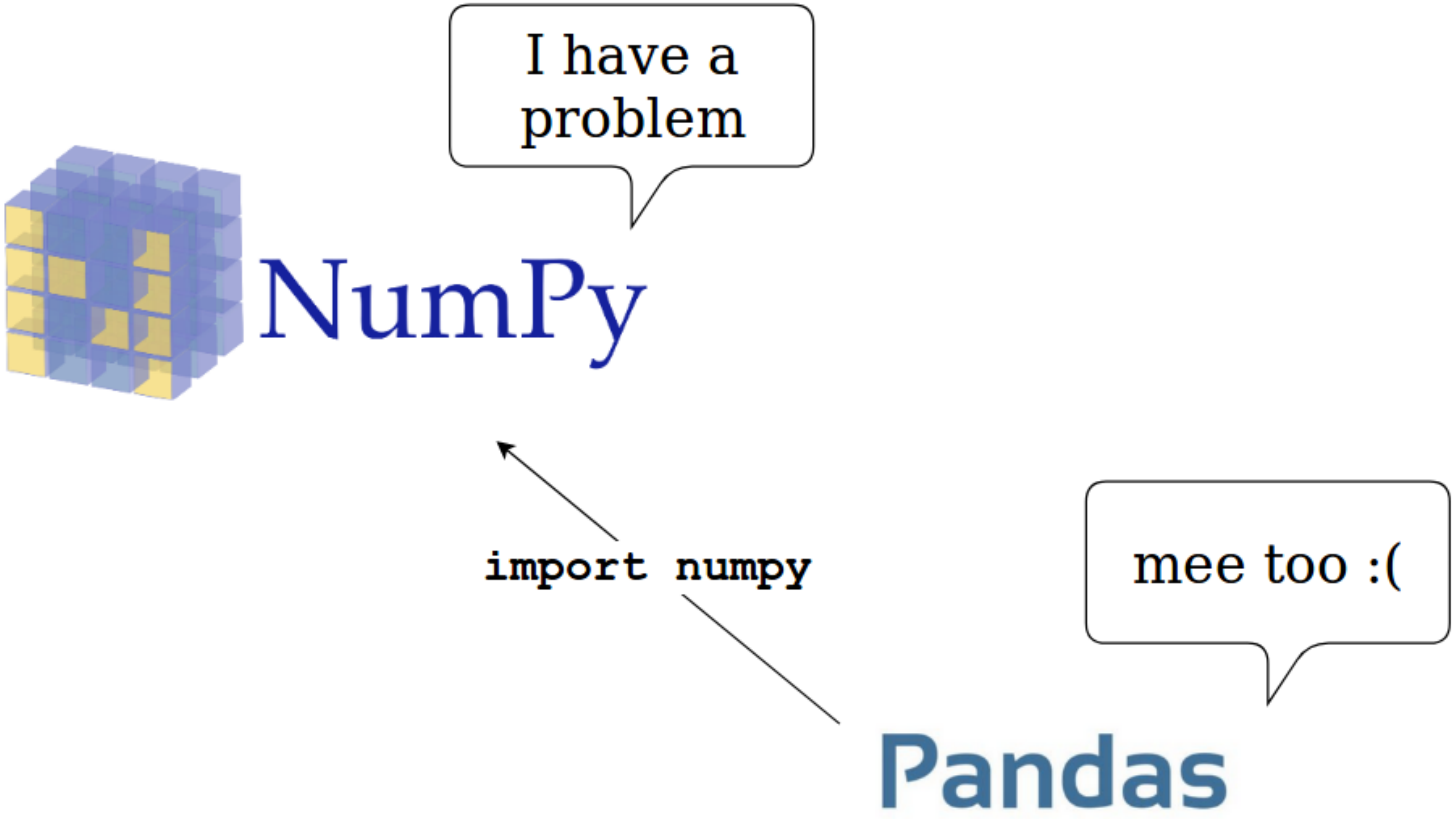
Cox survival regression

Interviews:

10 project maintainers



Are upstreams harmful?



I have a problem

NumPy

import numpy

mee too :(

Pandas

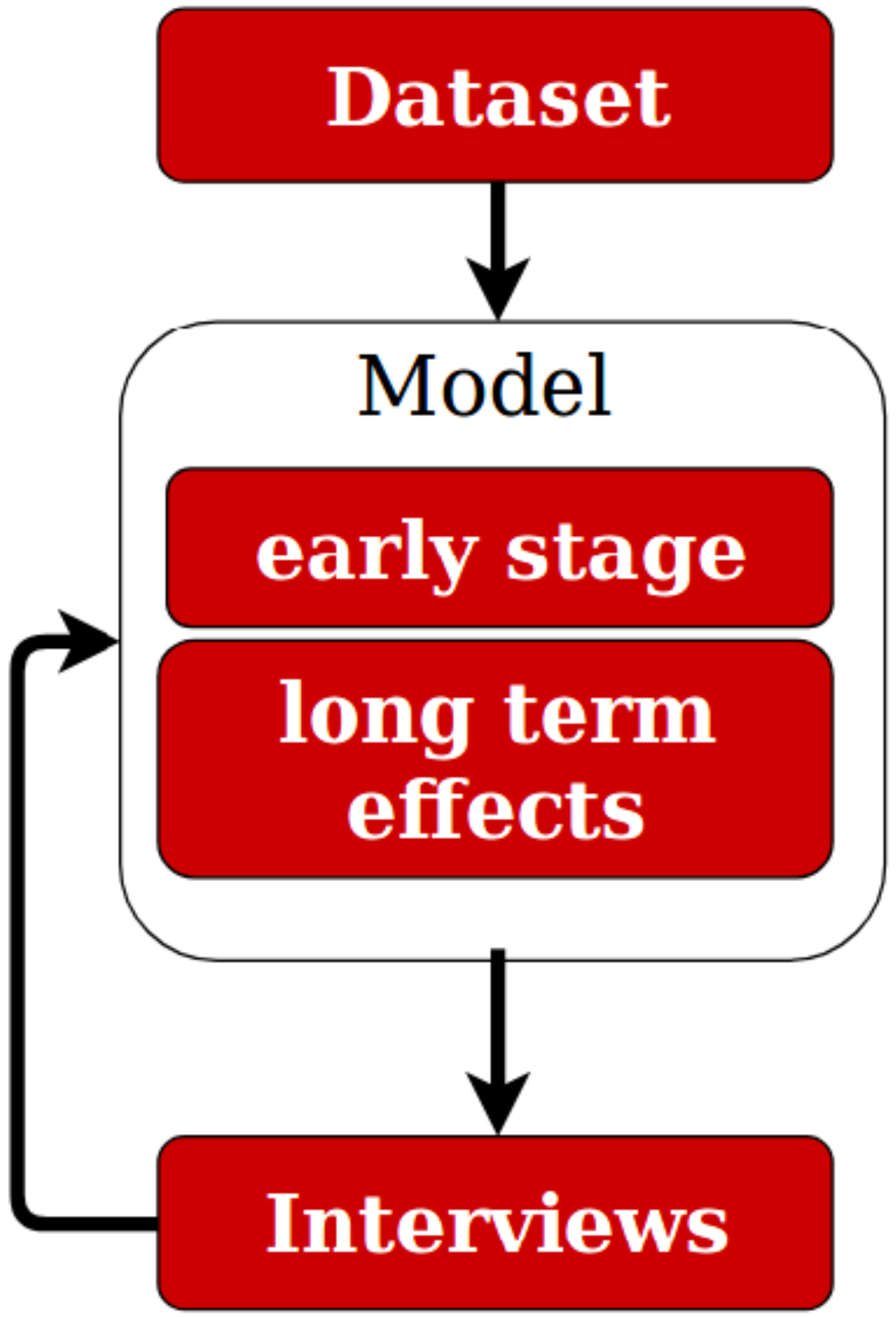
Upstreams are not always harmful

Feature: number of upstream projects

Early stage: **-25%** survival with every extra upstream
Long term: **+5%**

Interviews:

- conserve effort to reimplement dependency
- keep to the minimum, but not less
- added nonlinearity: no effect



Upstreams are not always harmful

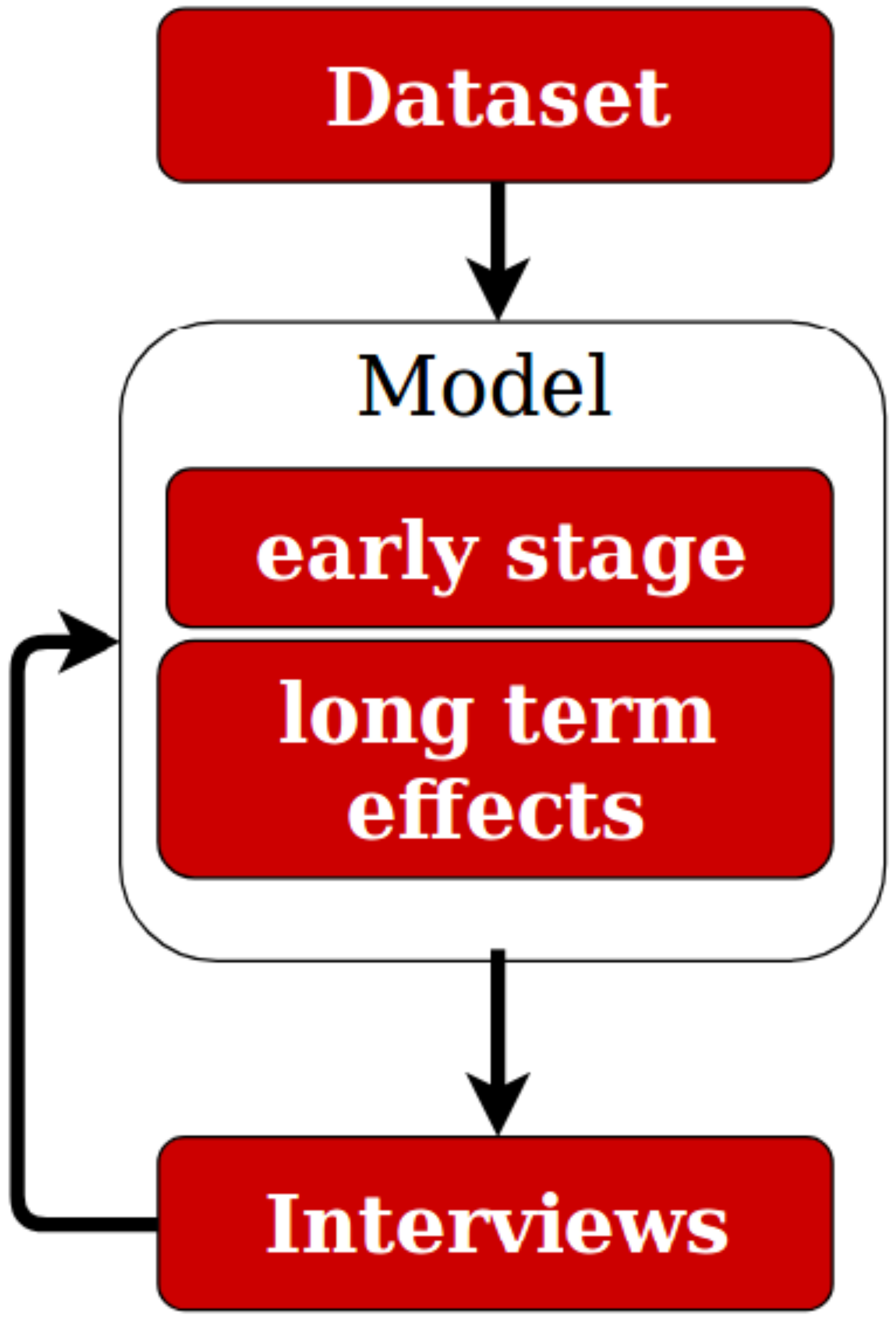
Feature: is any of the upstreams dormant?

Early stage: +31% to survival

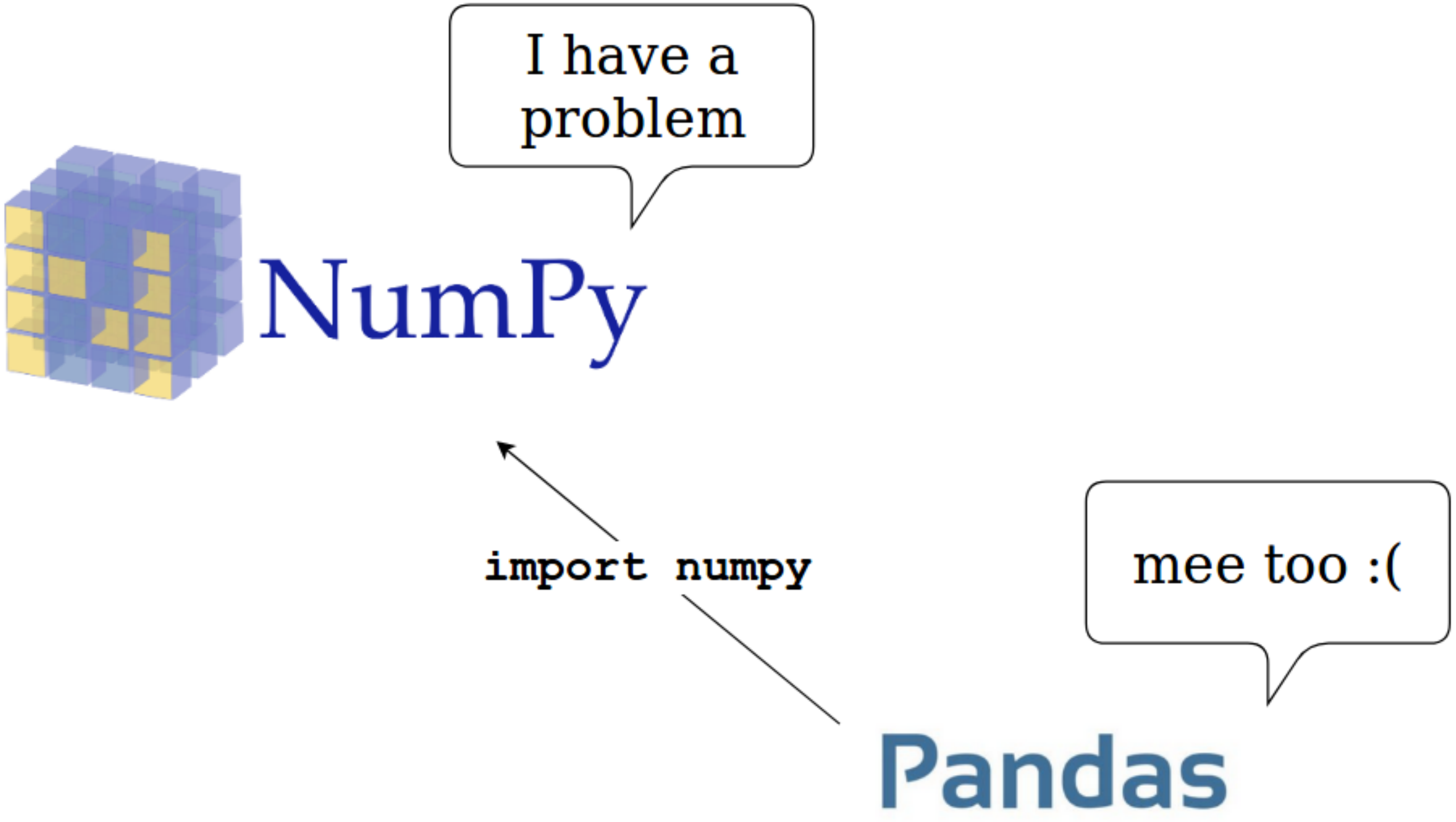
Long term: -11%

Interviews:

- feature complete projects (e.g., RFC standard) are dormant



Are downstreams helpful?



I have a problem

NumPy

import numpy

mee too :(

Pandas

Downstreams are helpful (long term)

Feature:

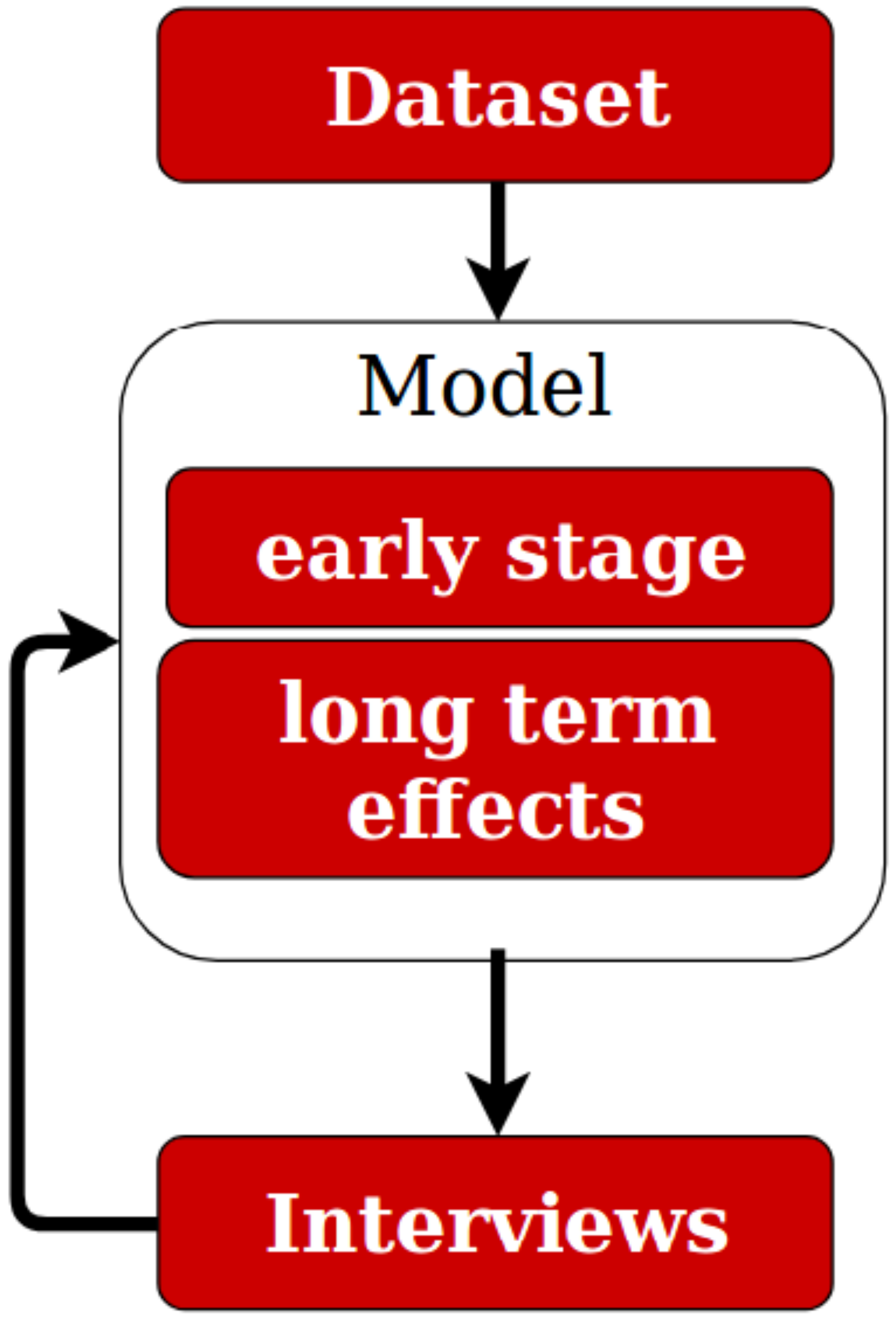
number of downstream projects

Early stage: **-60%** to survival

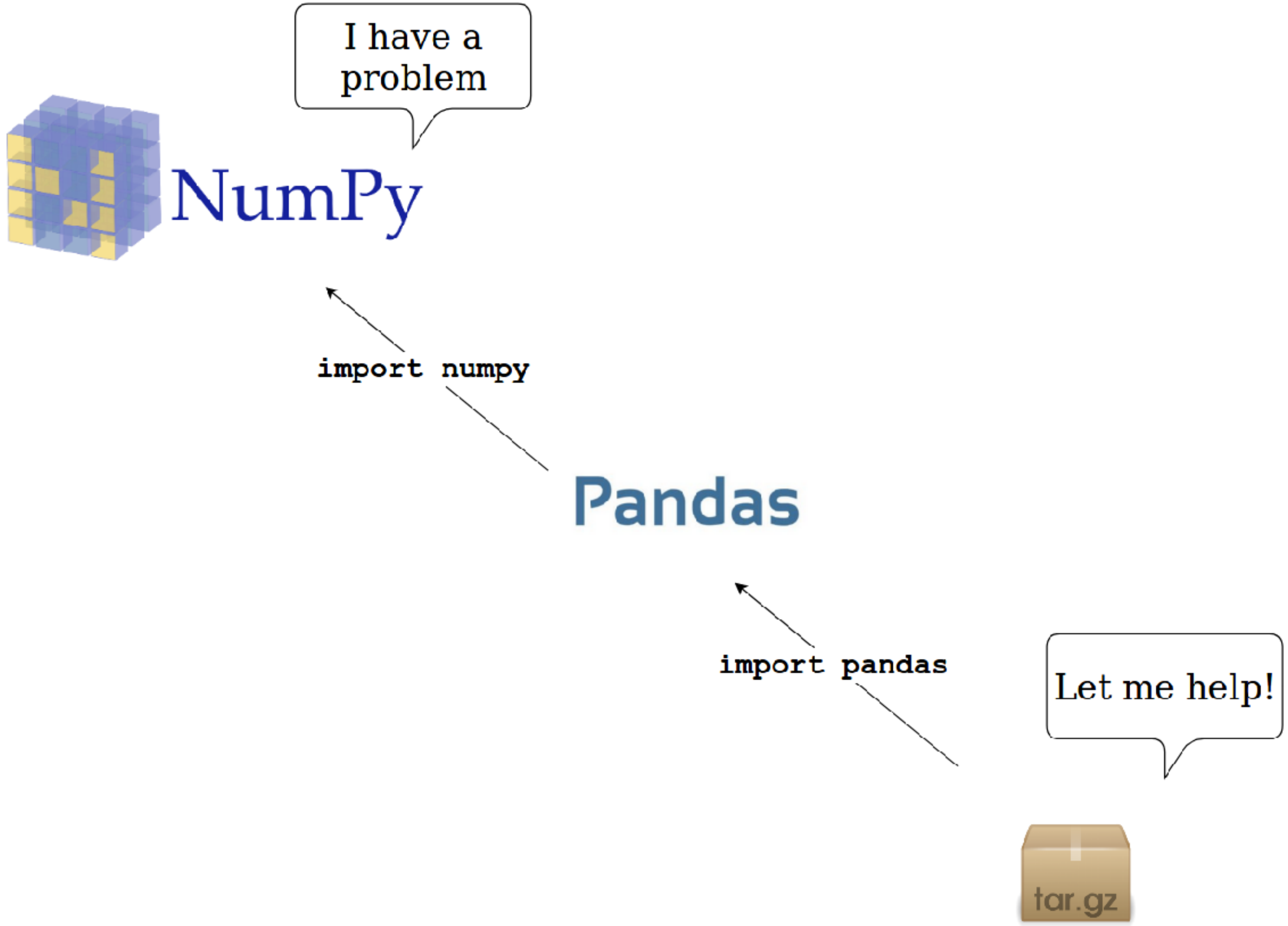
Long term: **+11%**

Interviews:

- contributors and free testers
- early stage: chip-off projects
- e.g., https://github.com/zopefoundation/Zope



Are transitive downstreams helpful?



Transitive downstreams are harmful

Feature: Katz centrality

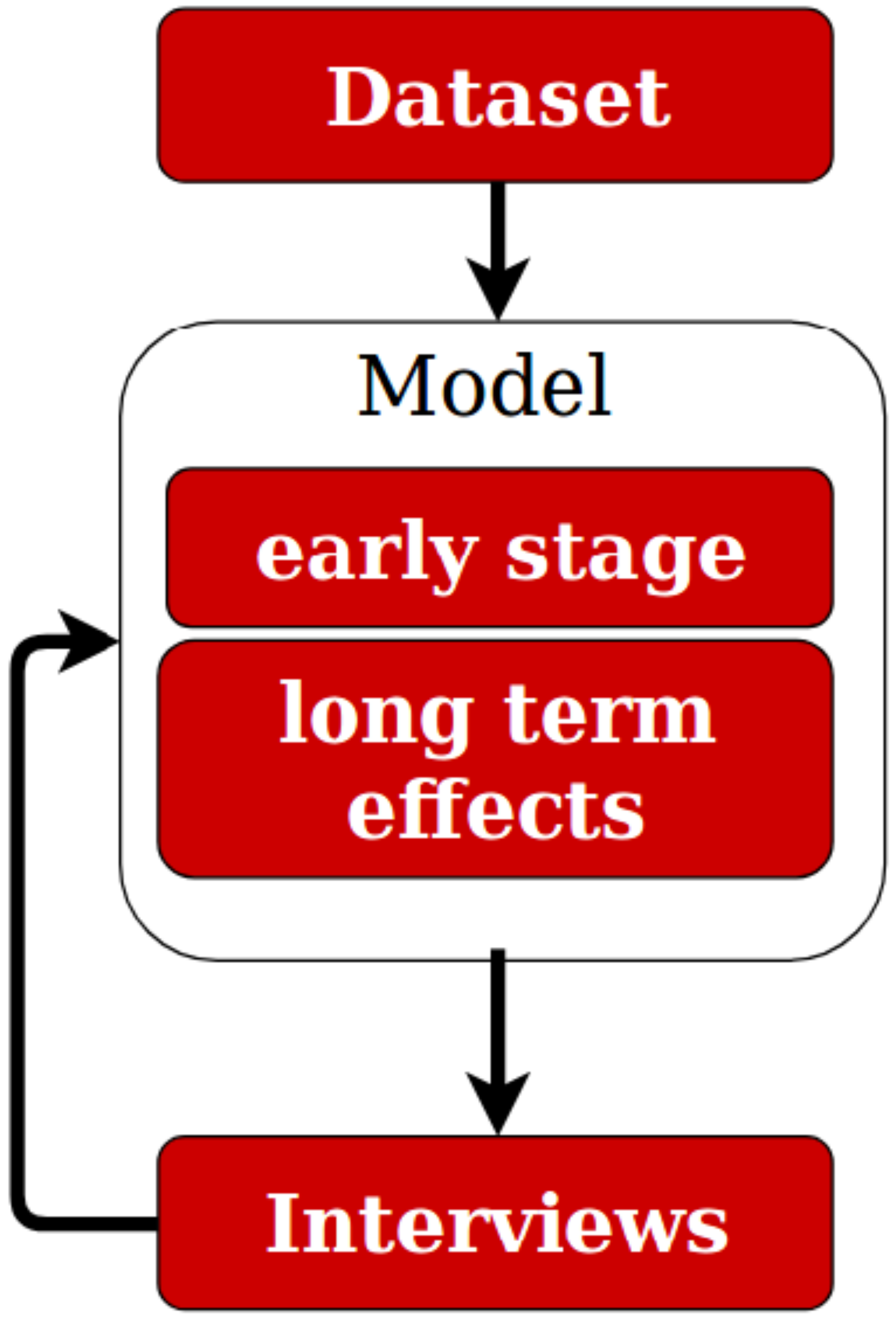
(discounted transitive dependencies)

Early stage: **-12%** to survival

Long term: **-27%**

Interviews:

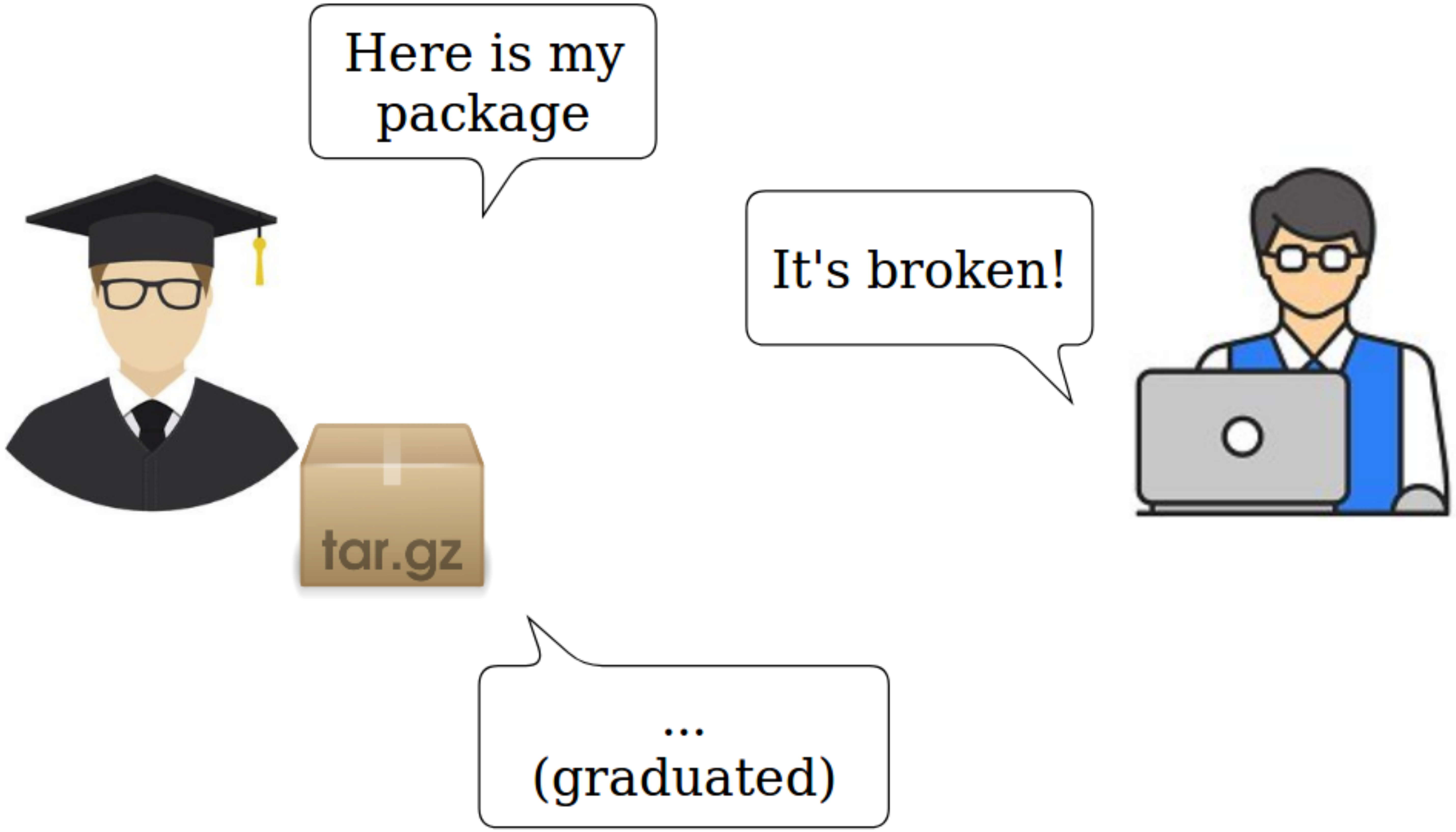
- less likely to fix
- just as likely to complain



Is support from large organizations helpful?

I have a problem
scikit learn
Let me help!
APERTURE LABORATORIES

Are academic projects less sustainable?



Academic involvement is helpful, long term

Feature:

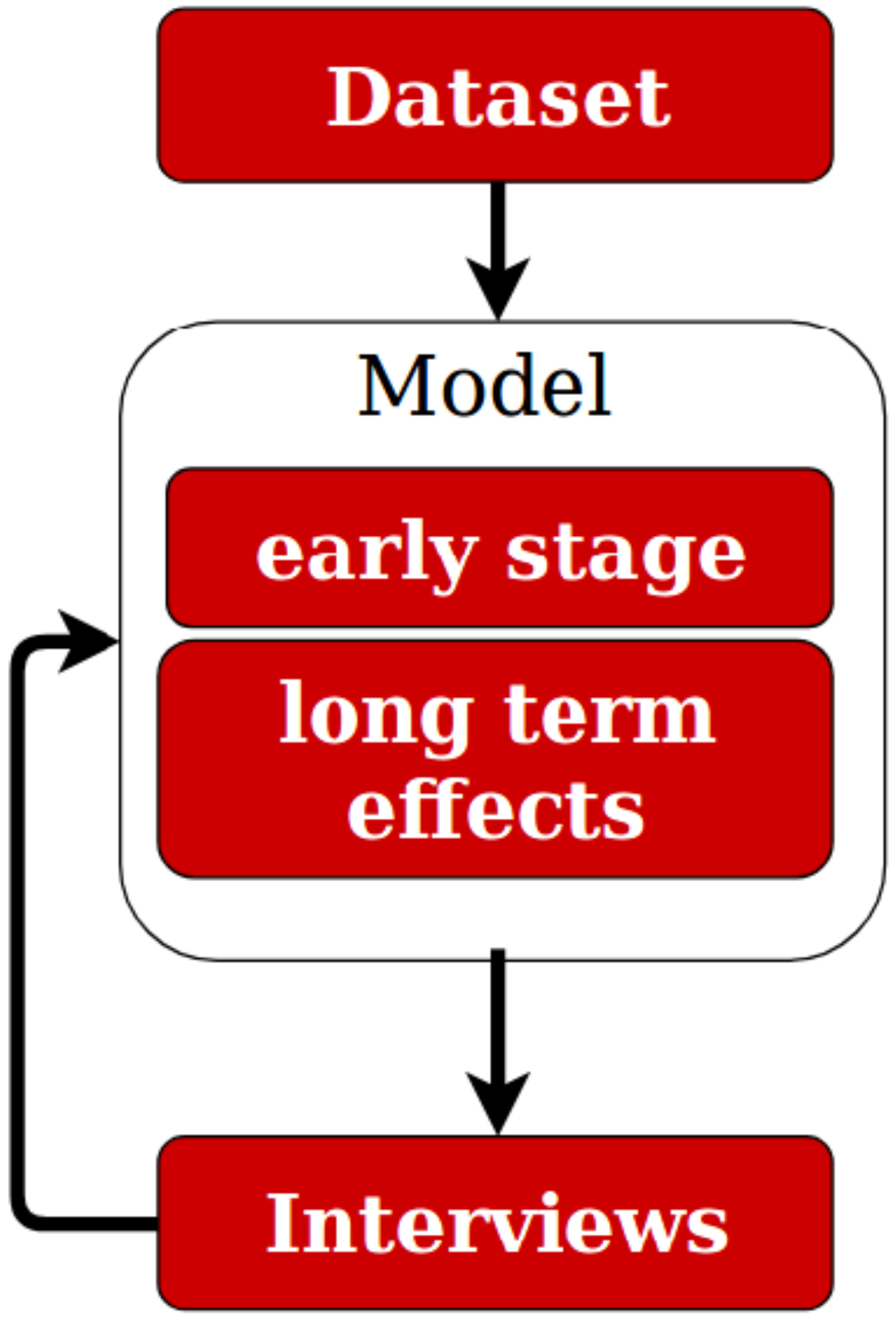
high academic involvement

Early stage: -8% to survival

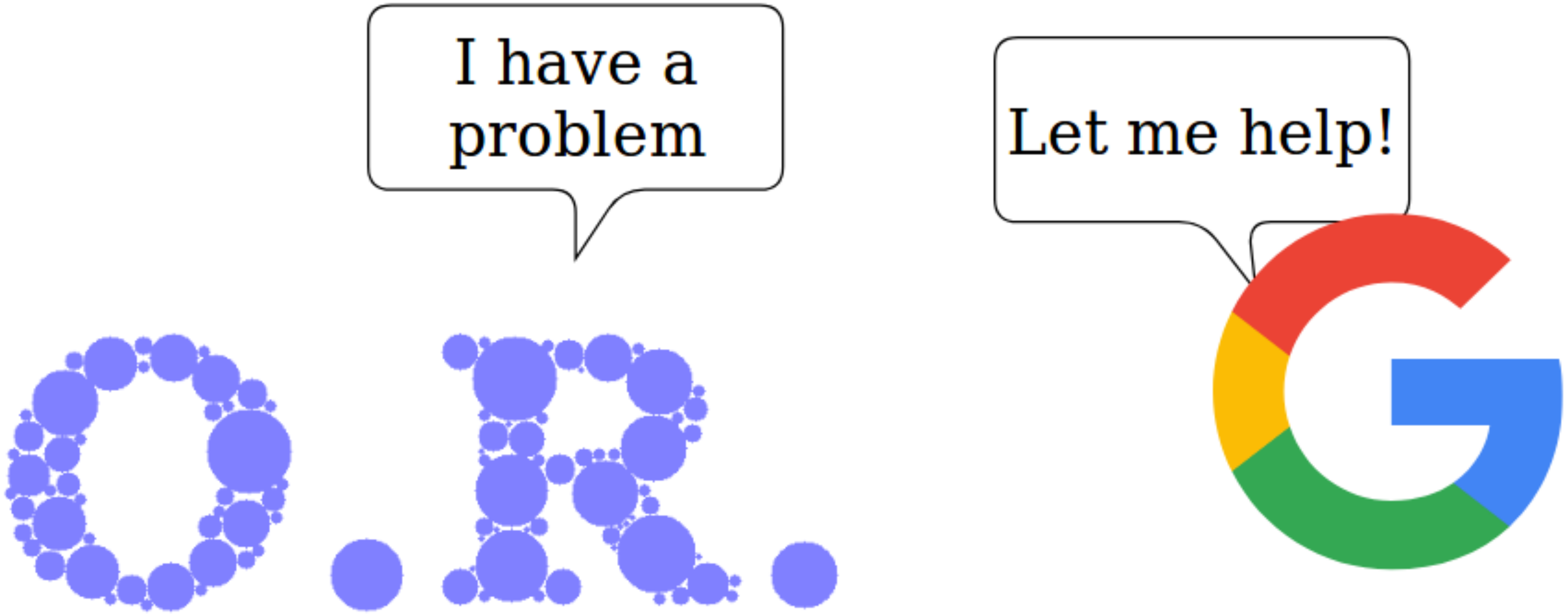
Long term: **+25%**

Interviews:

- projects supported by faculty
- continued funding is easier than initial



Are commercial projects more sustainable?



Commercial involvement is harmful

Feature:

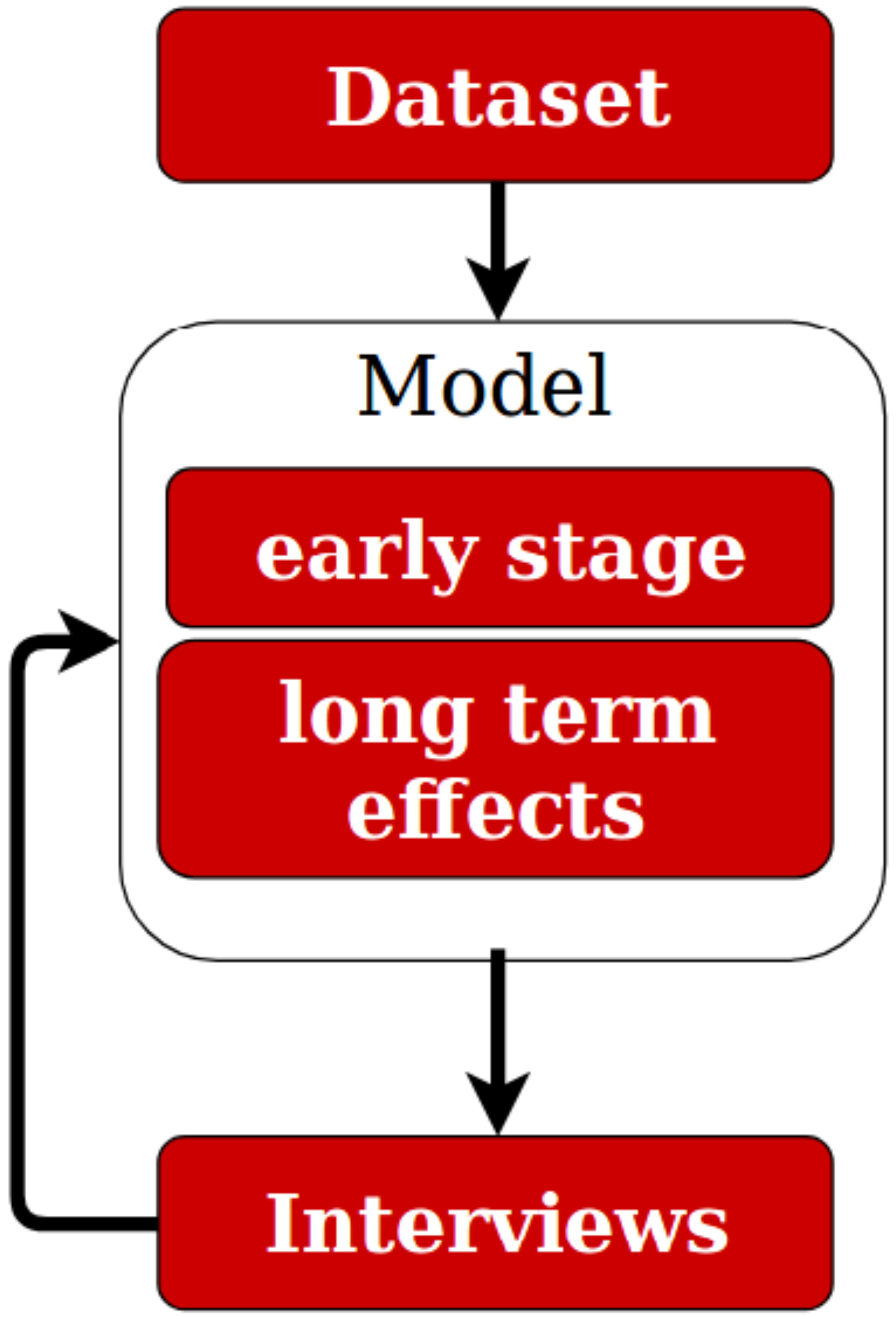
high commercial involvement

Early stage: -51% to survival

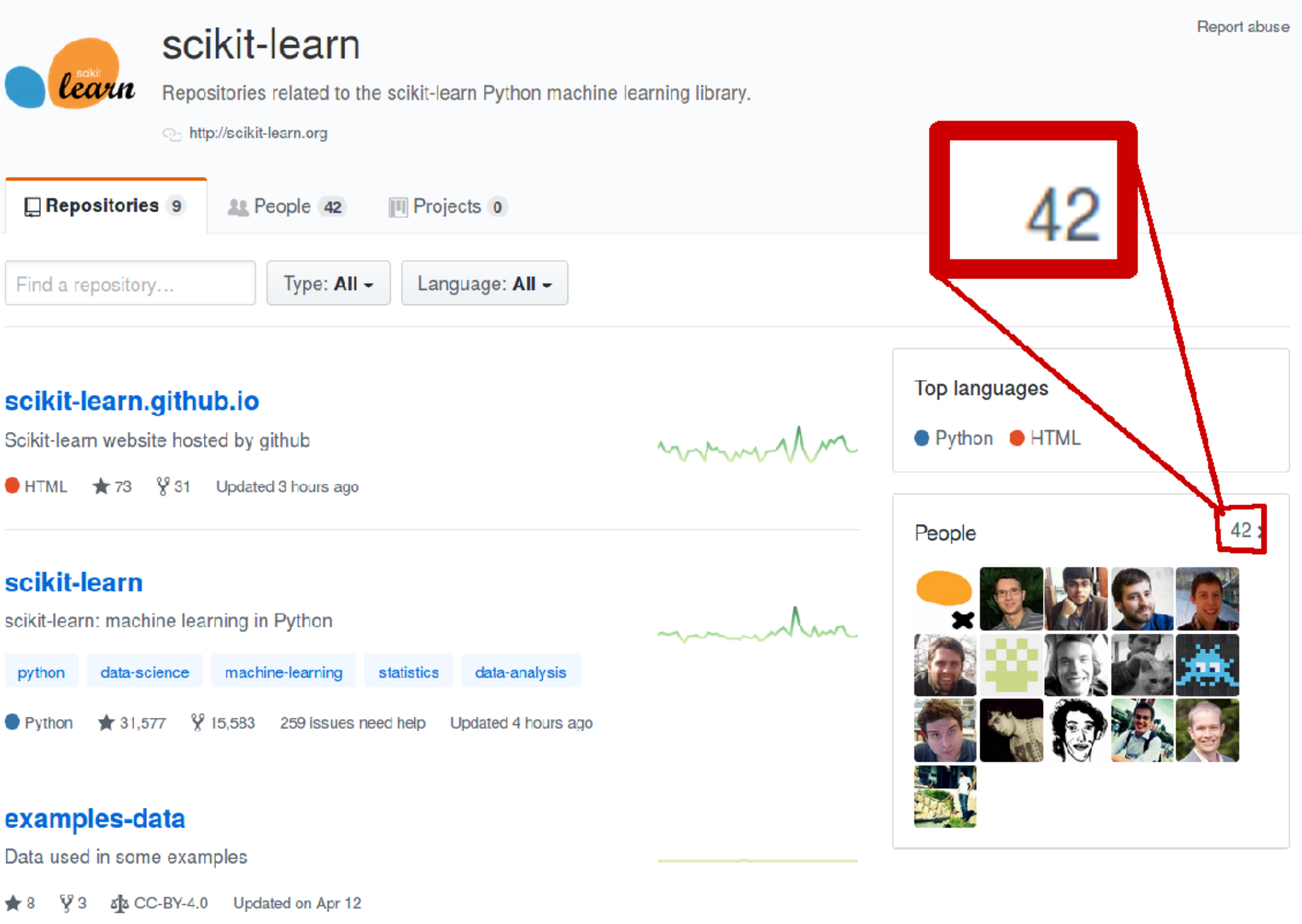
Long term: -15%

Interviews:

- companies bring more resources
- but they can withdraw anytime



Organizational accounts



Report abuse

scikit-learn

Repositories related to the scikit-learn Python machine learning library.

http://scikit-learn.org

Repositories 9 | People 42 | Projects 0

Find a repository… | Type: All | Language: All

scikit-learn.github.io

Scikit-learn website hosted by github

● HTML ★ 73 ⑂ 31 Updated 3 hours ago

scikit-learn

scikit-learn: machine learning in Python

python | data-science | machine-learning | statistics | data-analysis

● Python ★ 31,577 ⑂ 15,583 259 issues need help Updated 4 hours ago

examples-data

Data used in some examples

★ 8 ⑂ 3 CC-BY-4.0 Updated on Apr 12

Top languages

● Python ● HTML

People 42

Hosting under an organizational account is helpful

Feature:

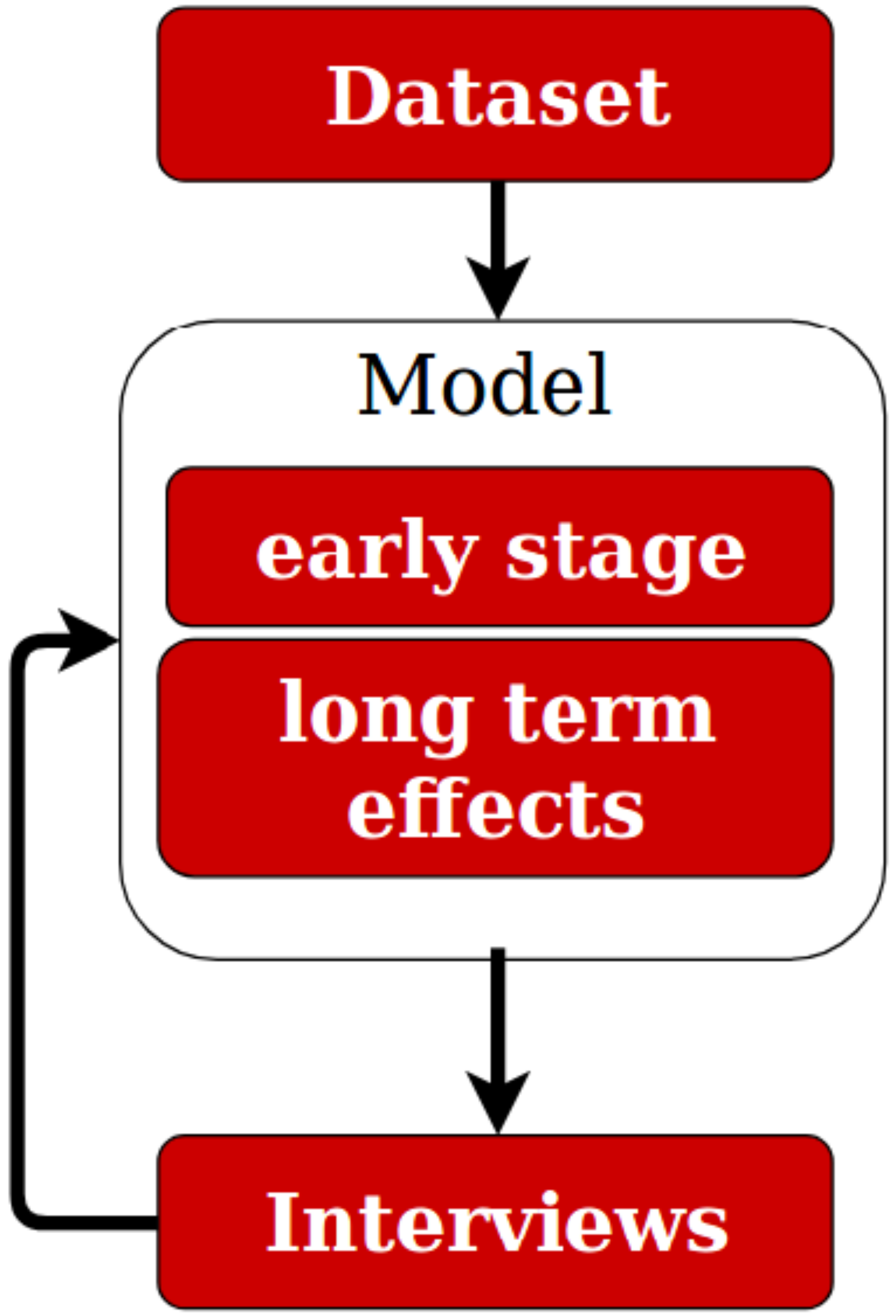
hosted under an org account on GitHub

Early stage: +45% to survival

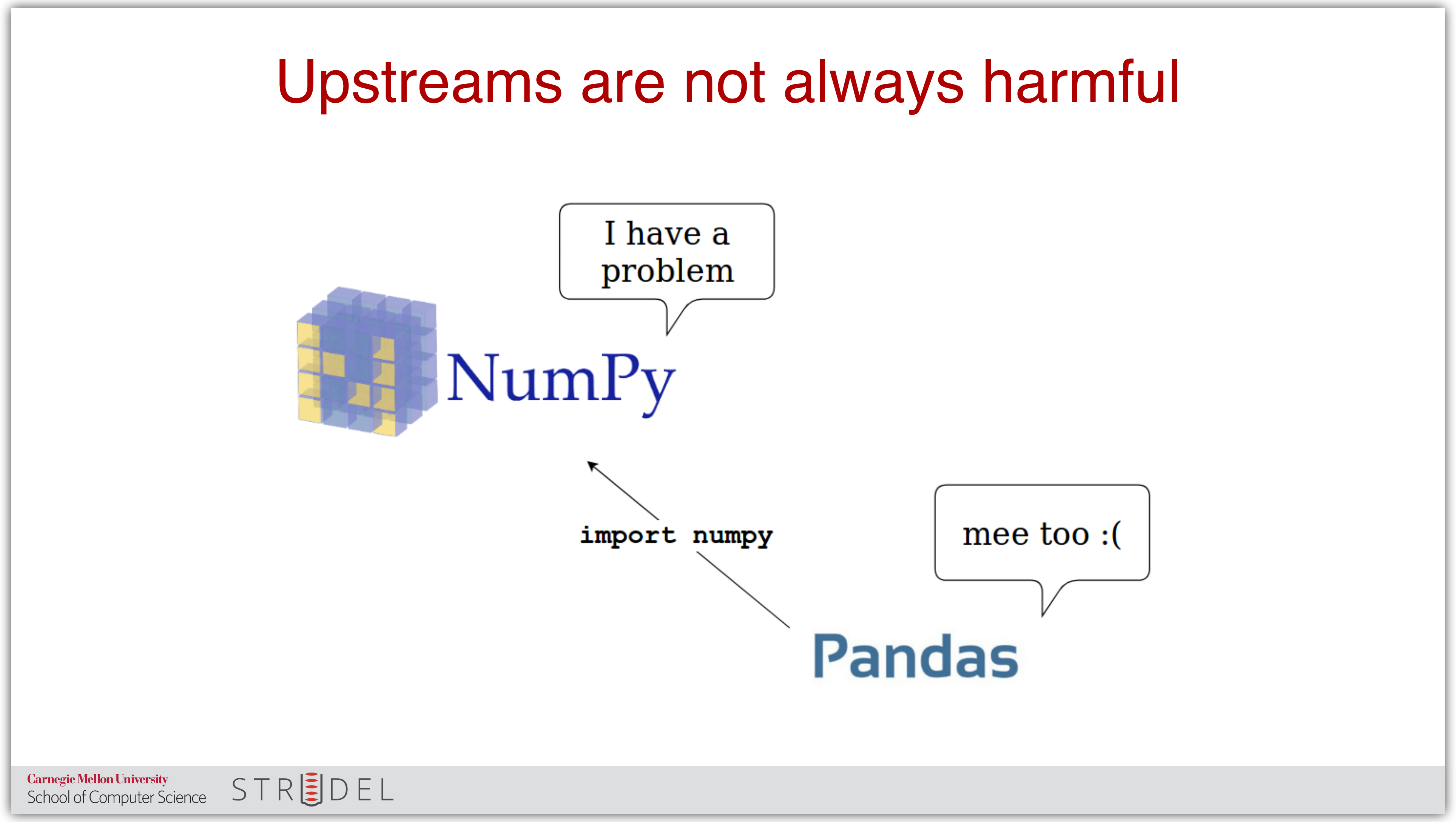
Long term: +23%

Interviews:

no strong opinion



External factors play an important role in the sustainability of open source projects



Upstreams are not always harmful

I have a problem

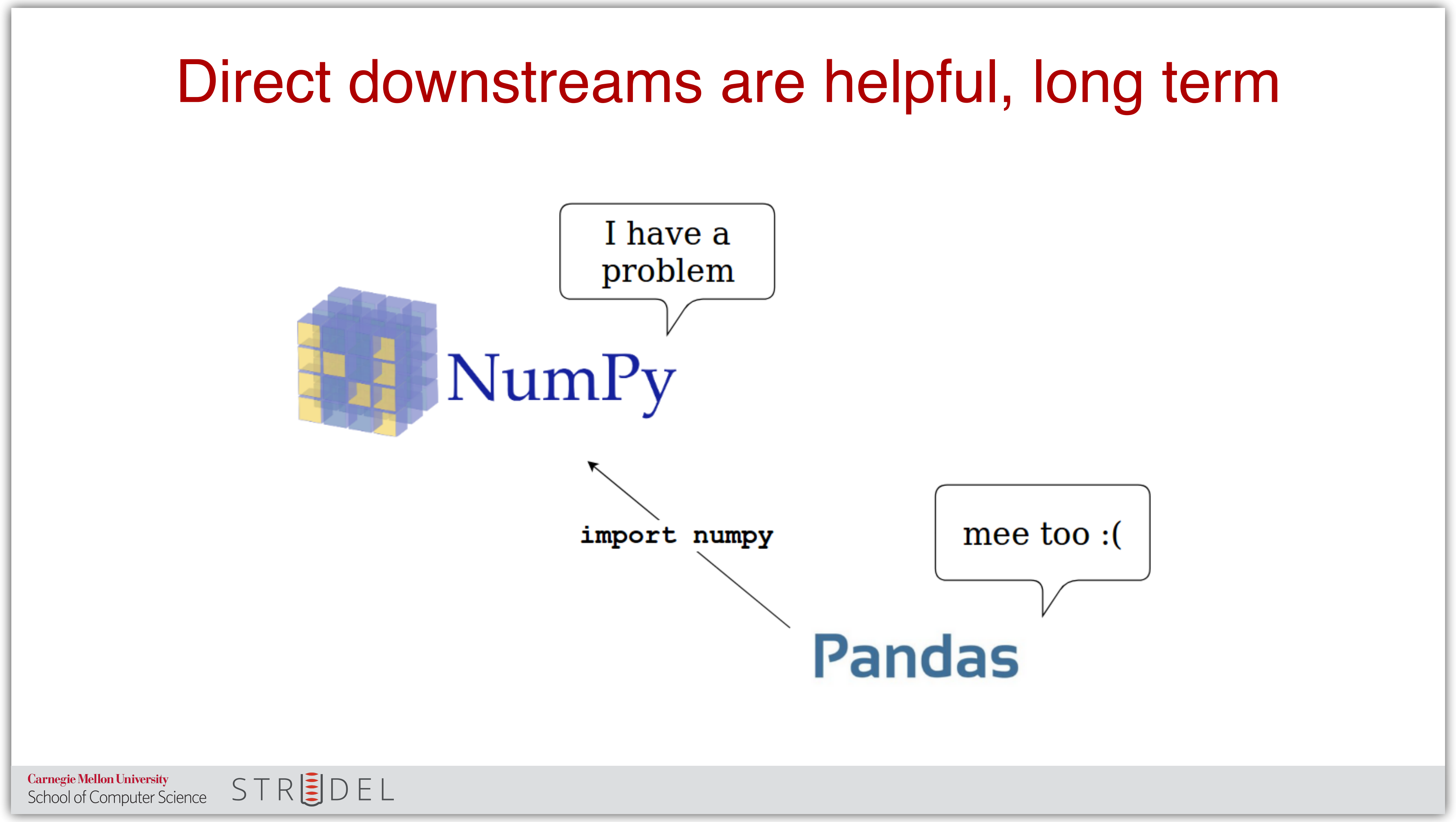
NumPy

import numpy

mee too :(

Pandas

Carnegie Mellon University
School of Computer Science STRUDEL



Direct downstreams are helpful, long term

I have a problem

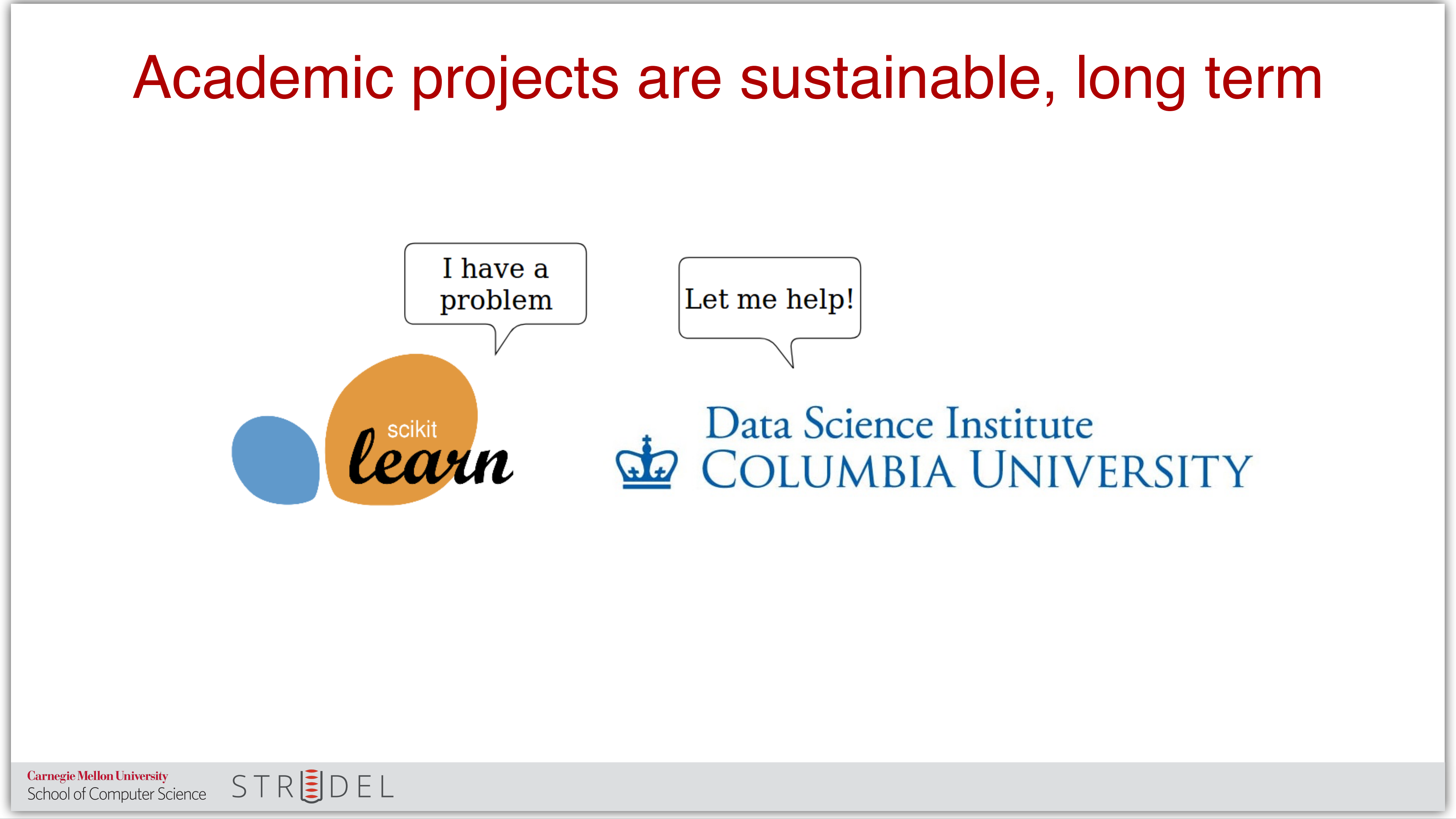
NumPy

import numpy

mee too :(

Pandas

Carnegie Mellon University
School of Computer Science STRUDEL



Academic projects are sustainable, long term

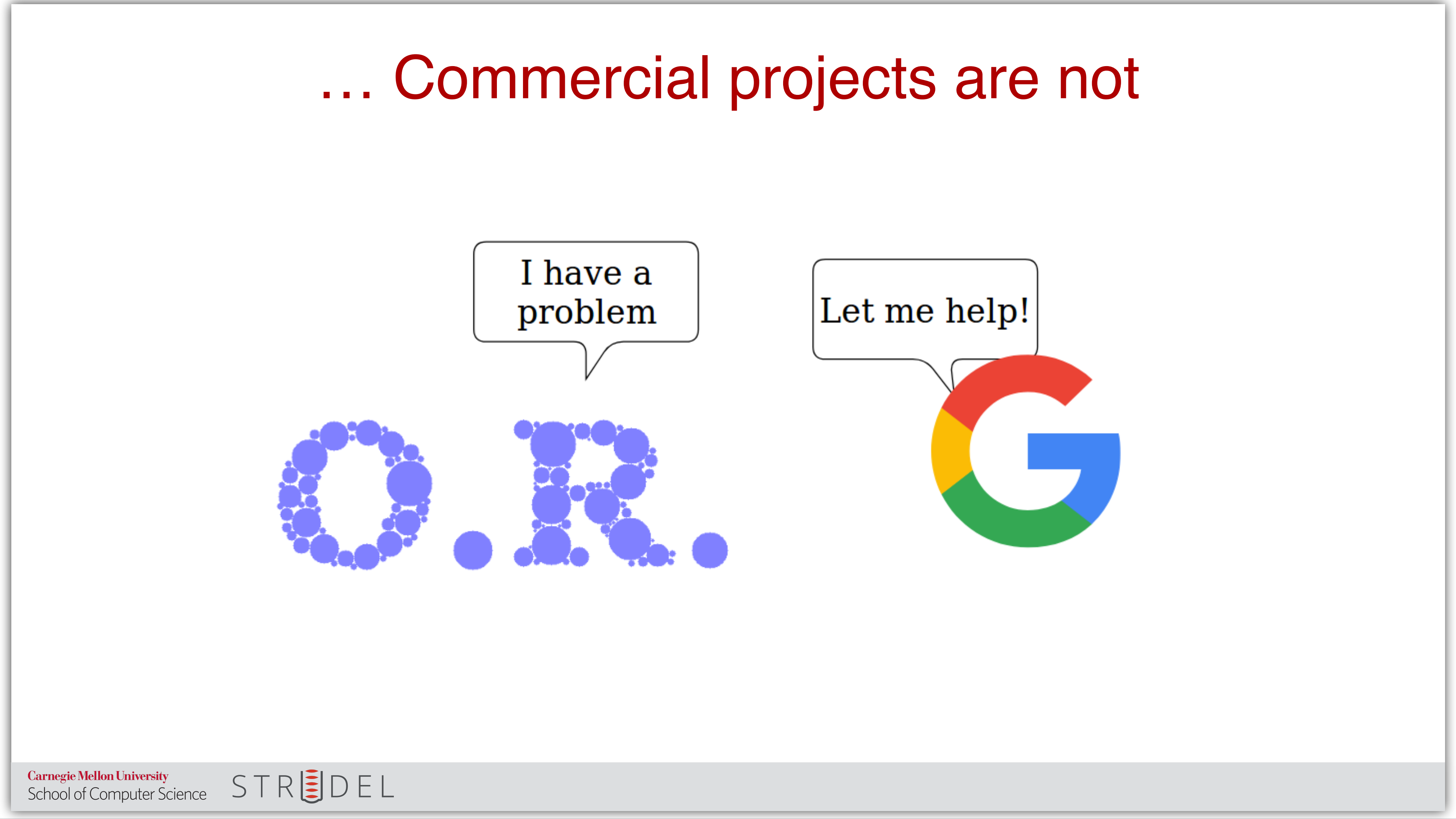
I have a problem

scikit learn

Let me help!

Data Science Institute
COLUMBIA UNIVERSITY

Carnegie Mellon University
School of Computer Science STRUDEL



… Commercial projects are not

I have a problem

O.R.

Let me help!

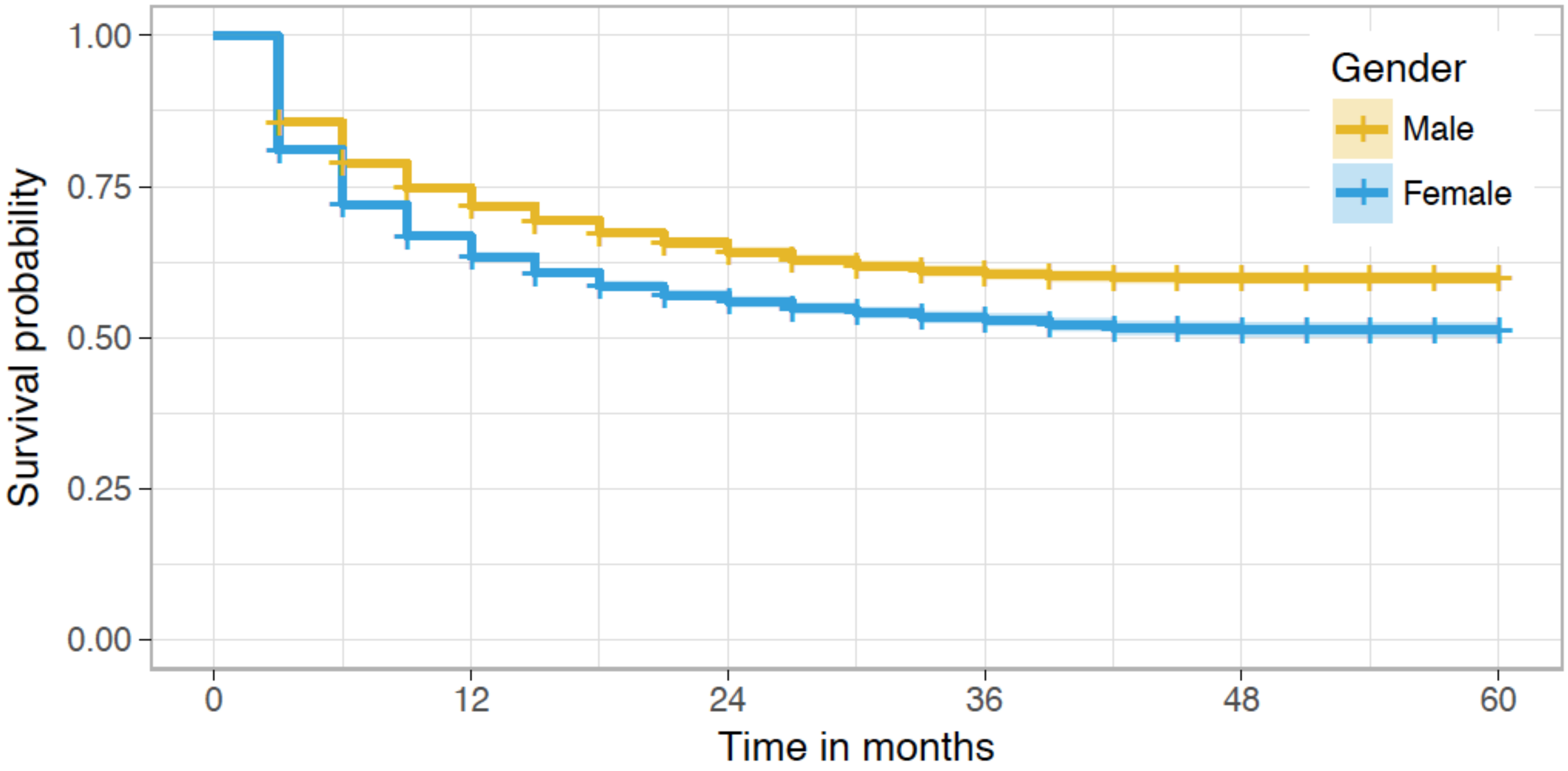
Carnegie Mellon University
School of Computer Science STRUDEL

Why do open source contributors disengage?

[Qiu et al. ICSE 2019]

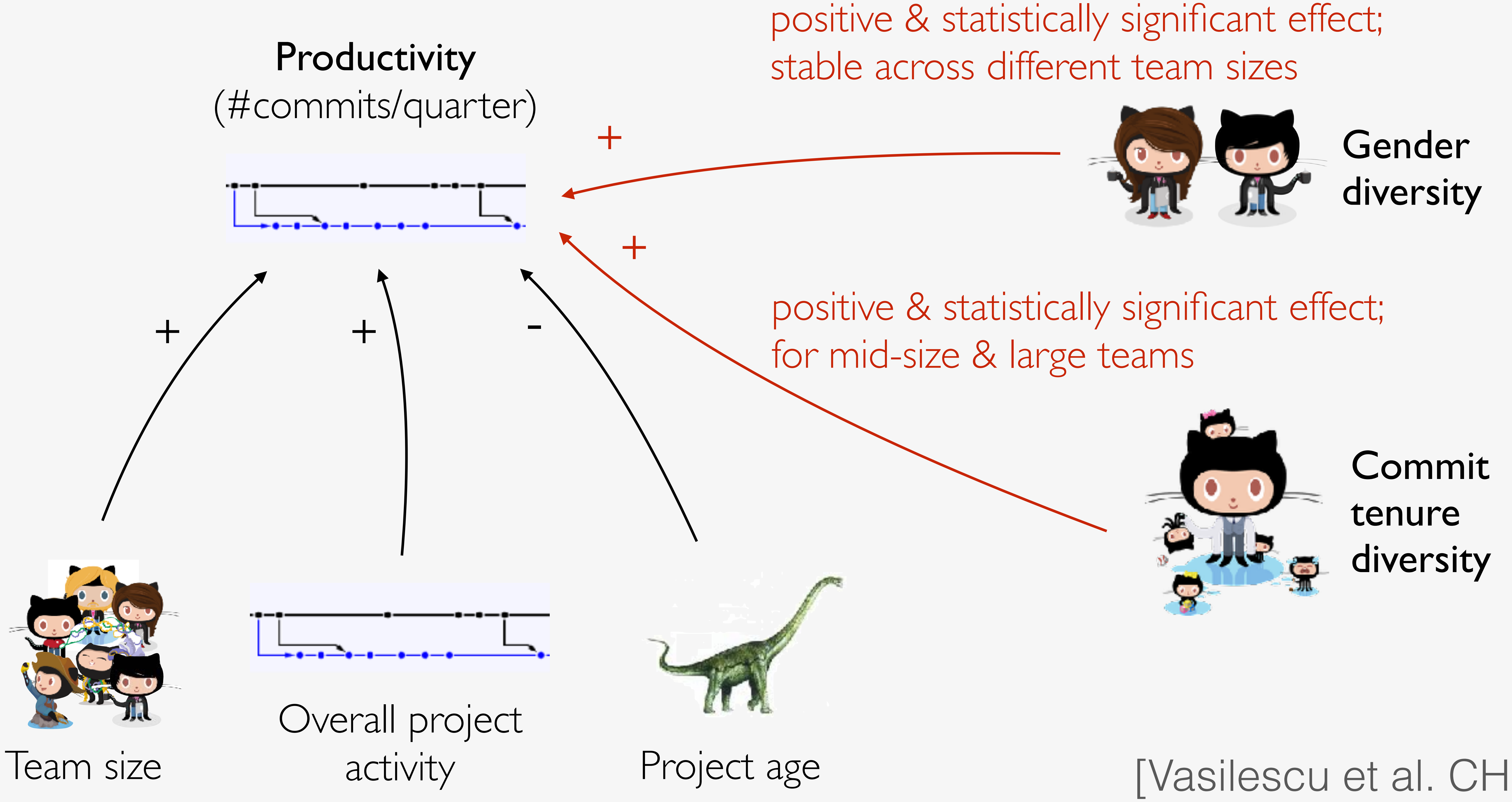
On GitHub, women disengage earlier than men

- After one year ca. 70% of men are still contributing to GitHub projects but only ca 60% of women



On GitHub, women disengage earlier than men

Aside: Other variables held fixed, more gender / tenure diverse teams are more productive than less diverse ones.



Productivity
(#commits/quarter)

positive & statistically significant effect;
stable across different team sizes

+

Gender diversity

+

positive & statistically significant effect;
for mid-size & large teams

Commit tenure diversity

+

+

-

Team size

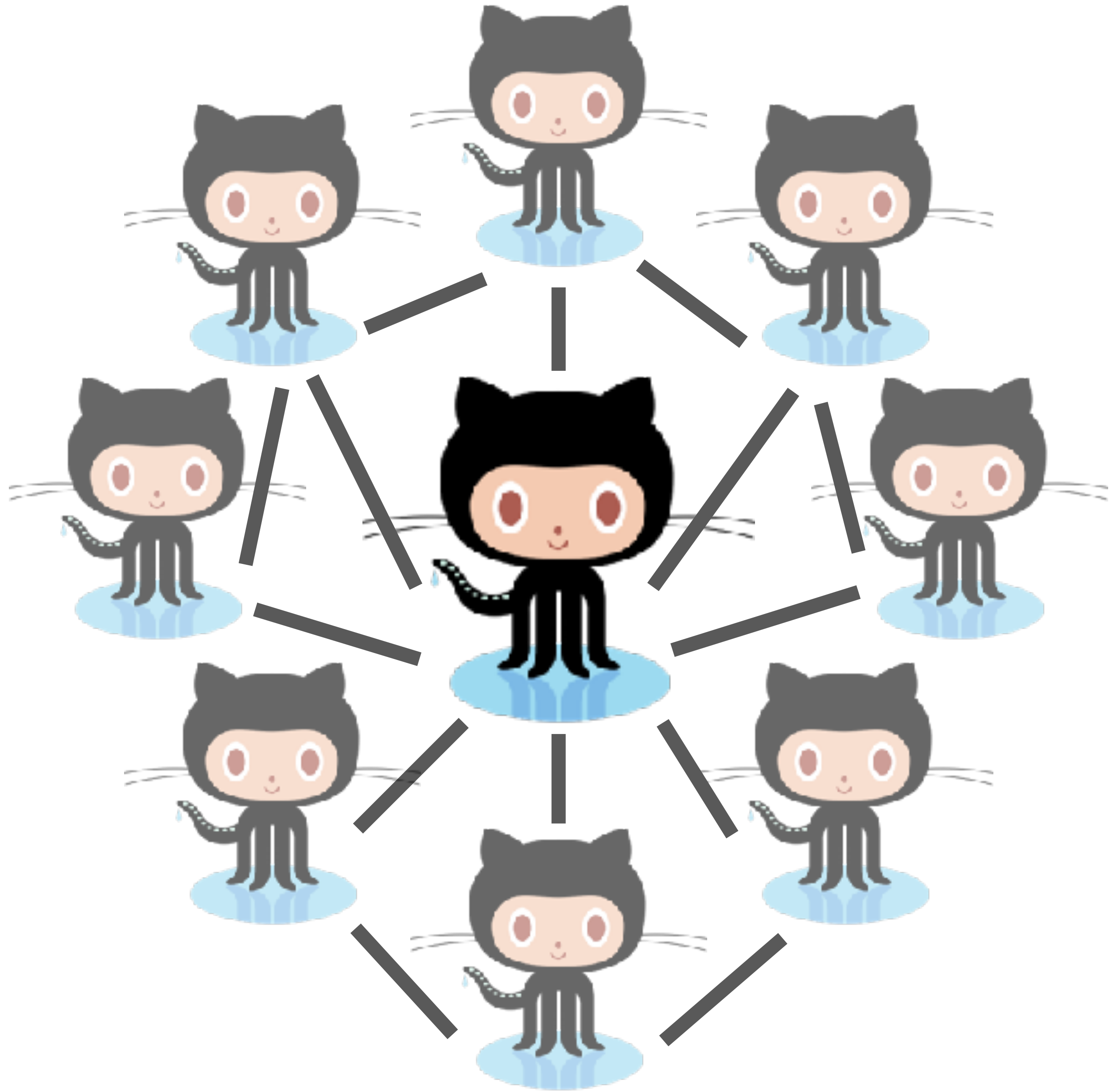
Overall project activity

Project age

[Vasilescu et al. CHI 2015]

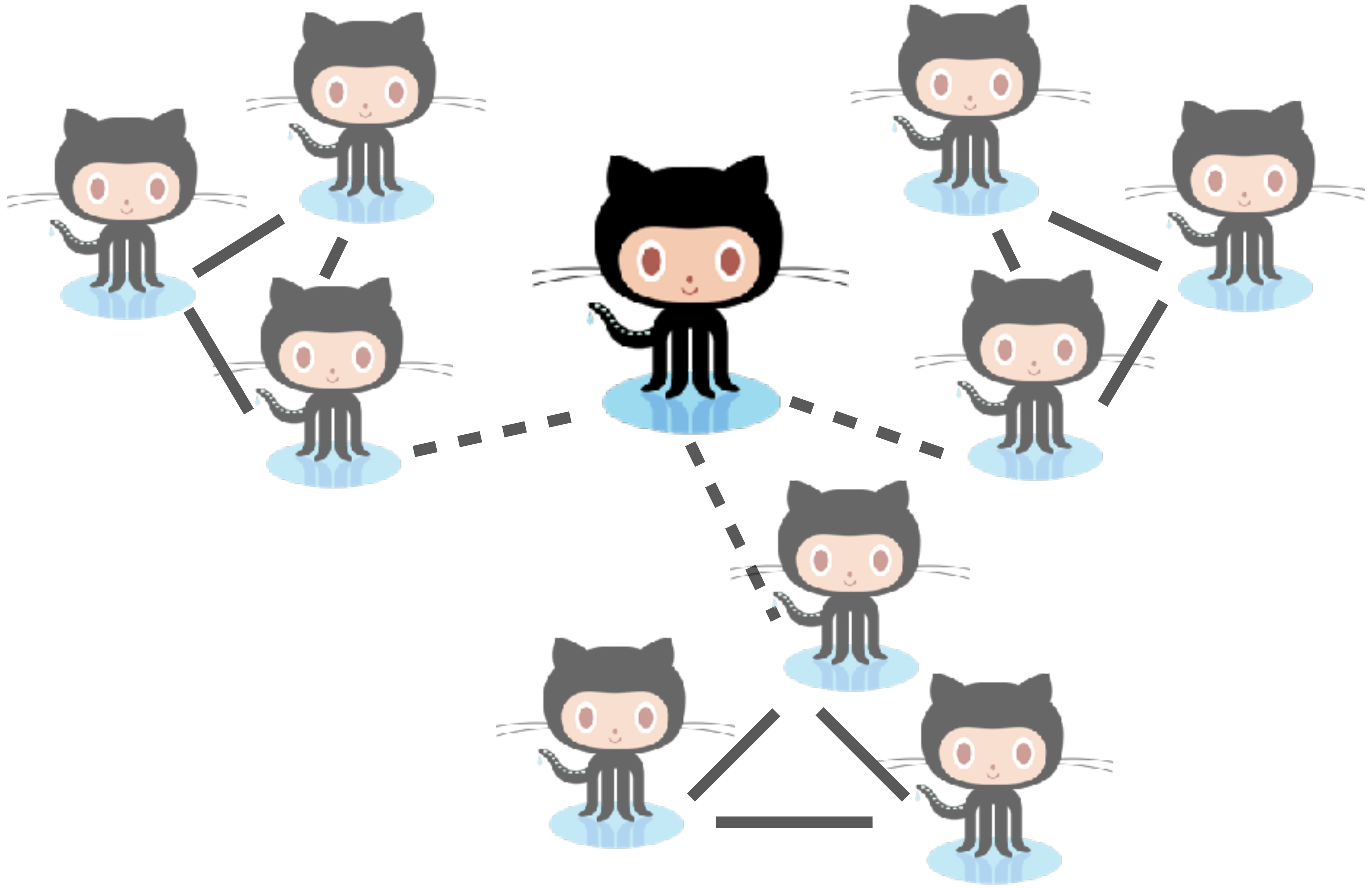
Social capital is the set of benefits individuals can gain from their social connections and social structures

Bonding social capital: benefiting from network closure



Willingness to continue

Bridging social capital: benefiting from a brokerage position

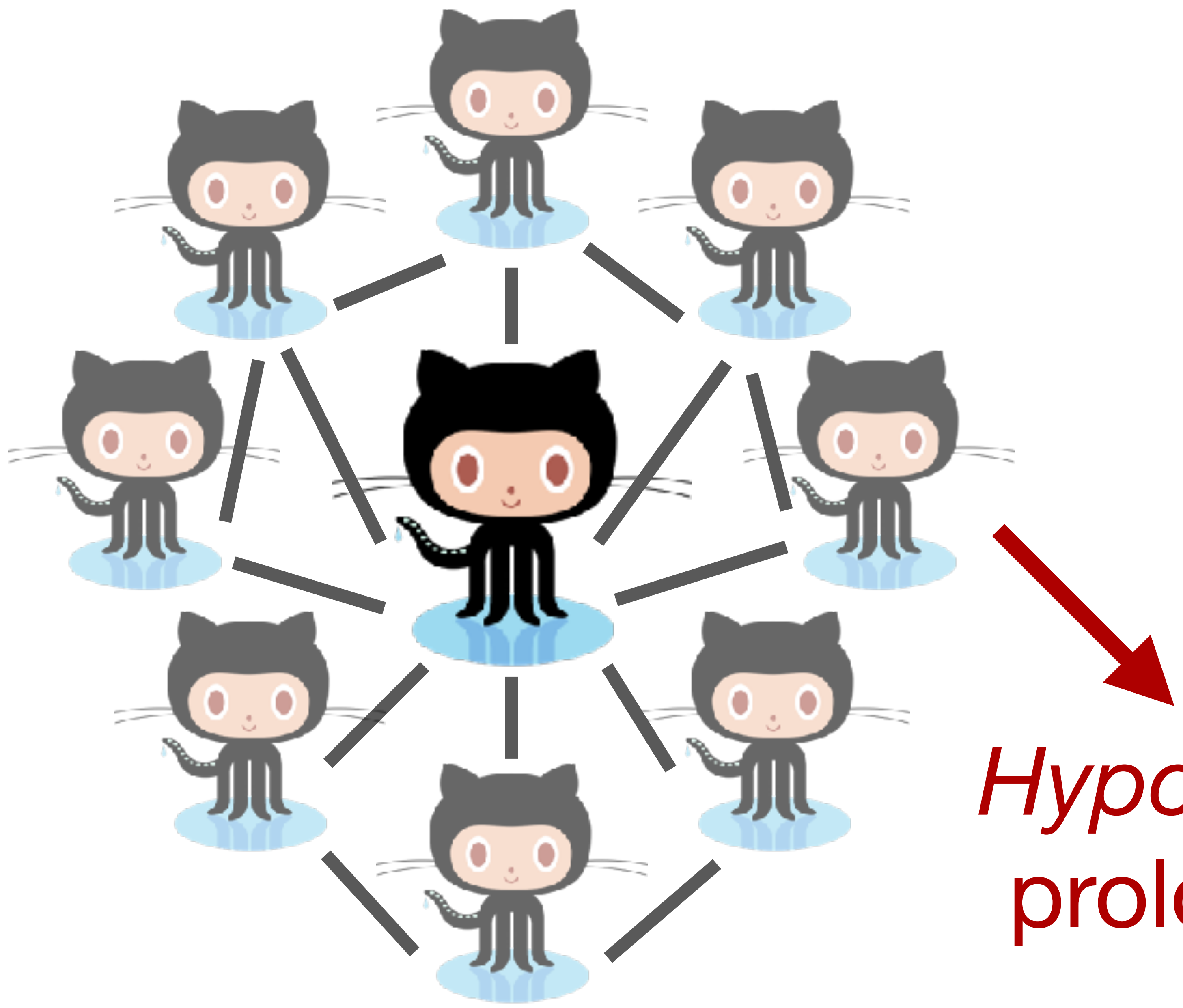


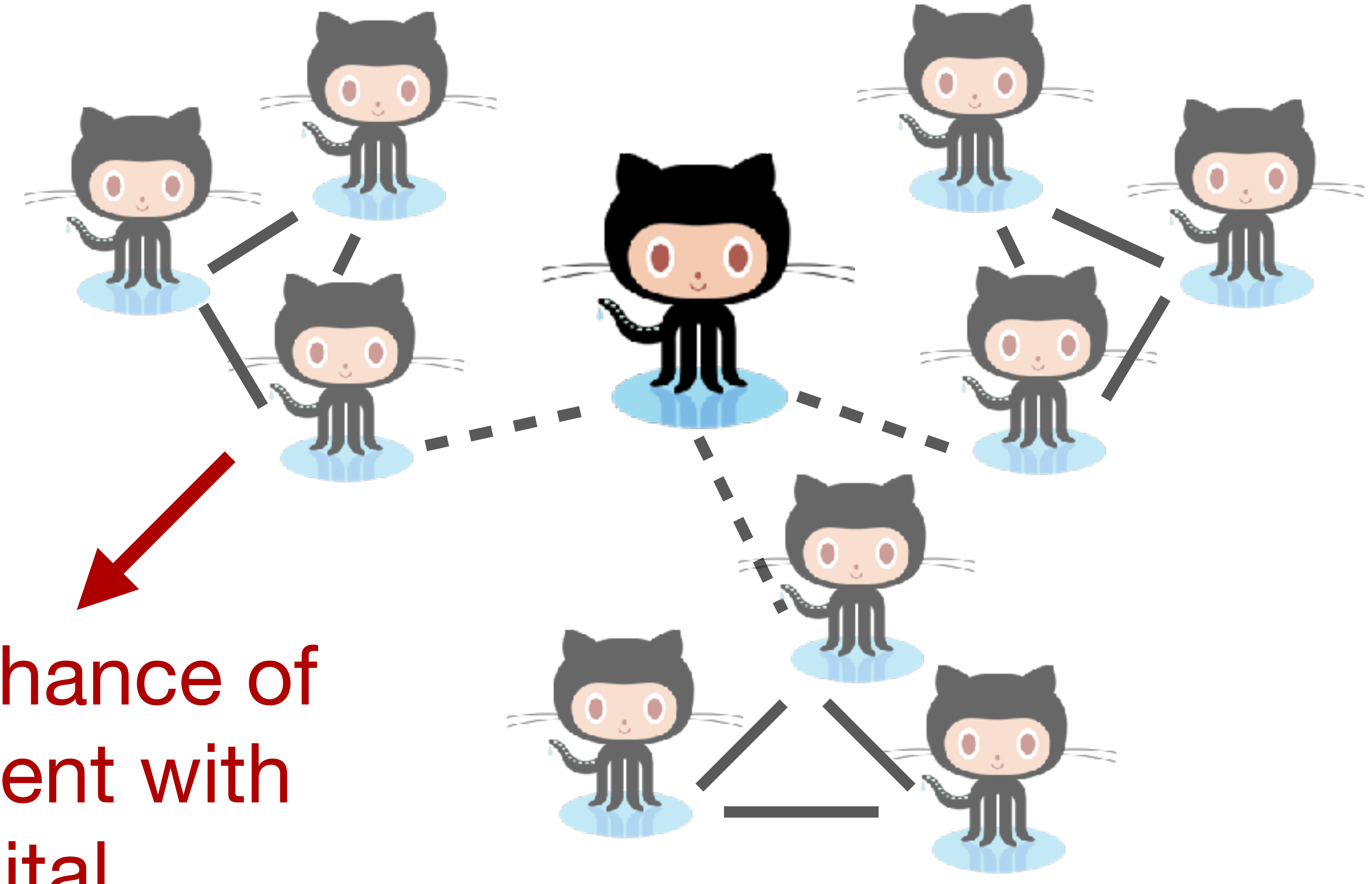
Opportunity to continue

Social capital is the set of benefits individuals can gain from their social connections and social structures

Bonding **social capital: benefiting from network closure**

Bridging **social capital: benefiting from a brokerage position**





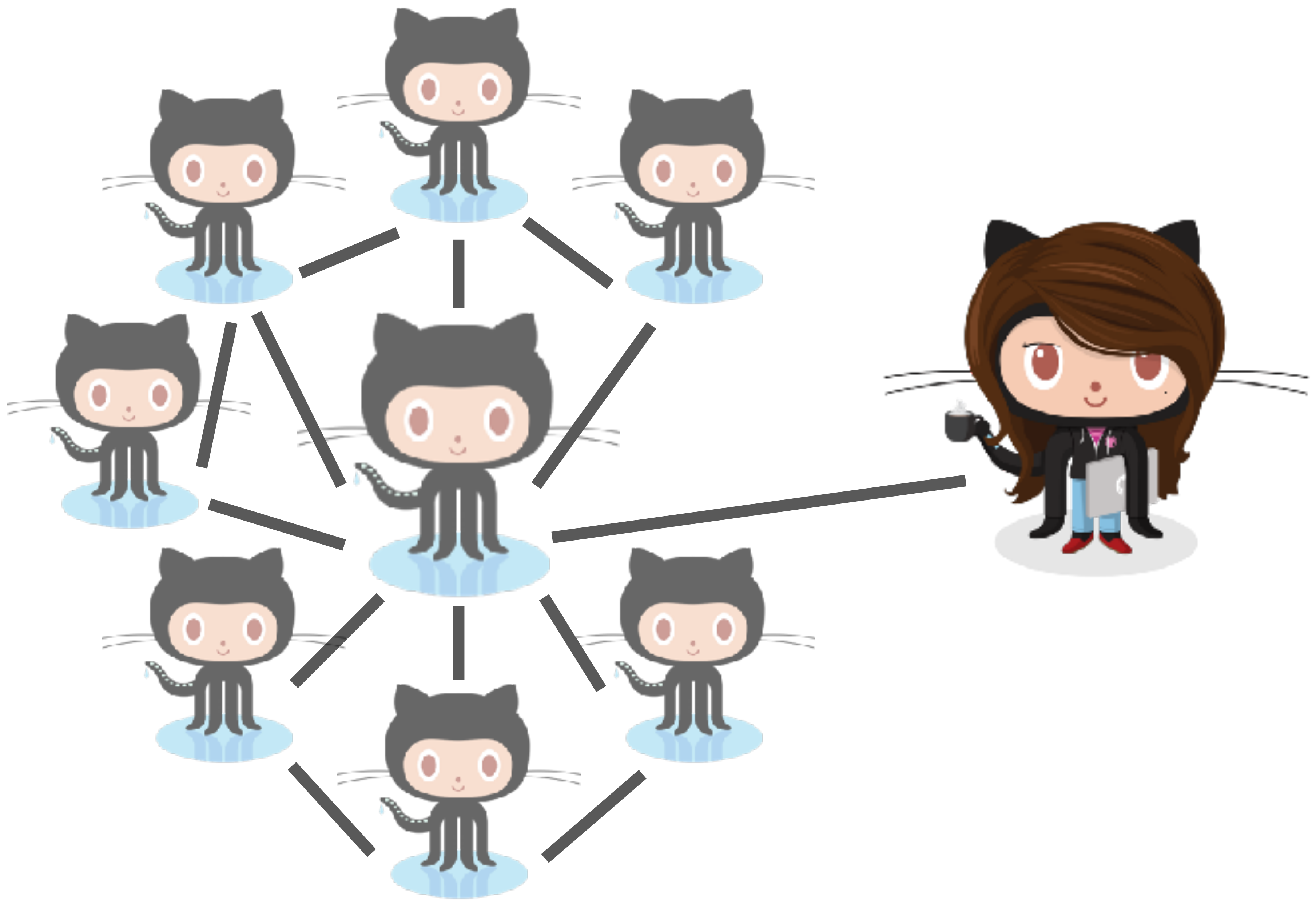
Hypothesis: Higher chance of prolonged engagement with more social capital.

Willingness to continue

Opportunity to continue

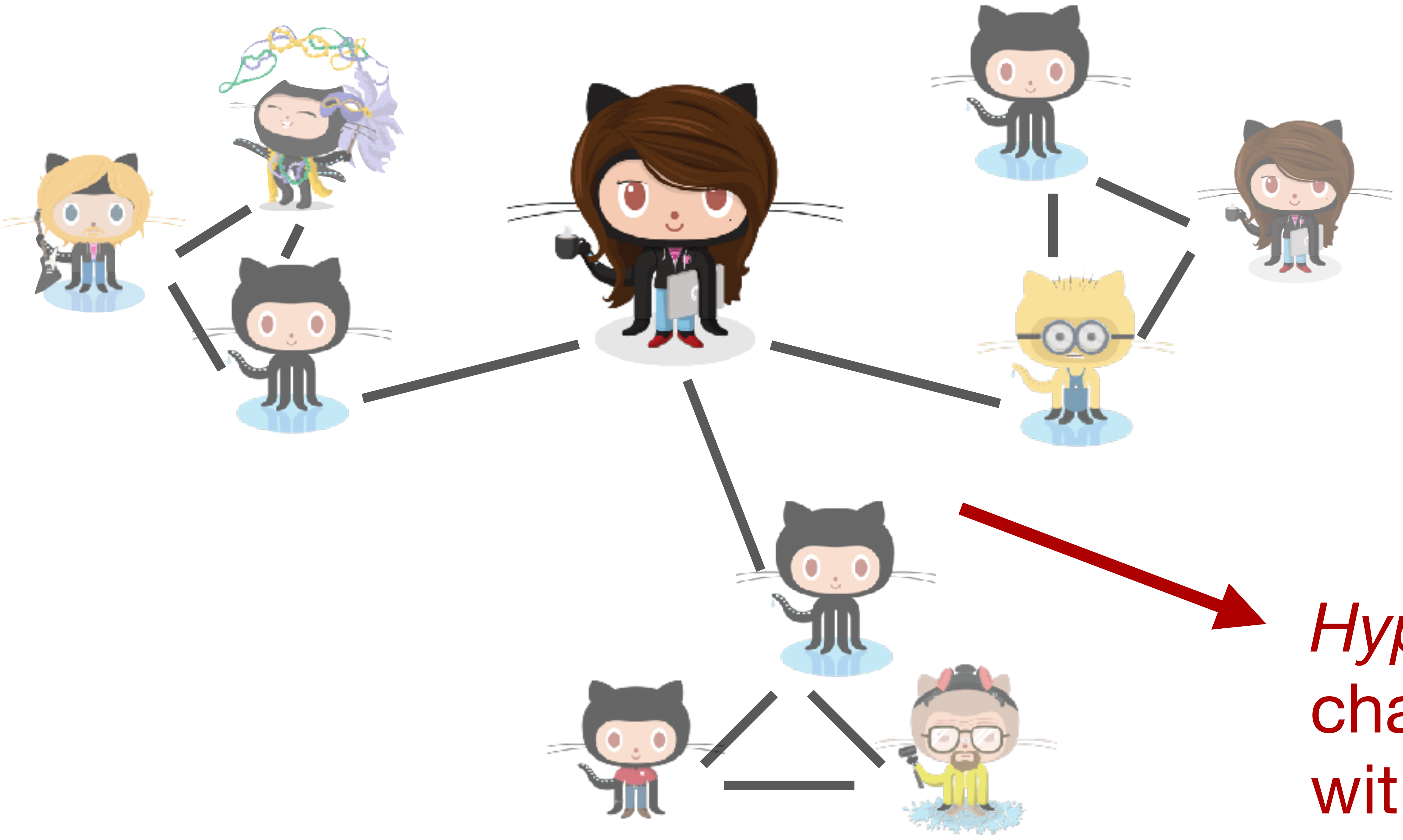
Network closure is likely to divide actors into insiders and outsiders

Cohesive networks might foster discrimination and exclusion



Since underrepresented, women tend to be outsiders, therefore at a disadvantage

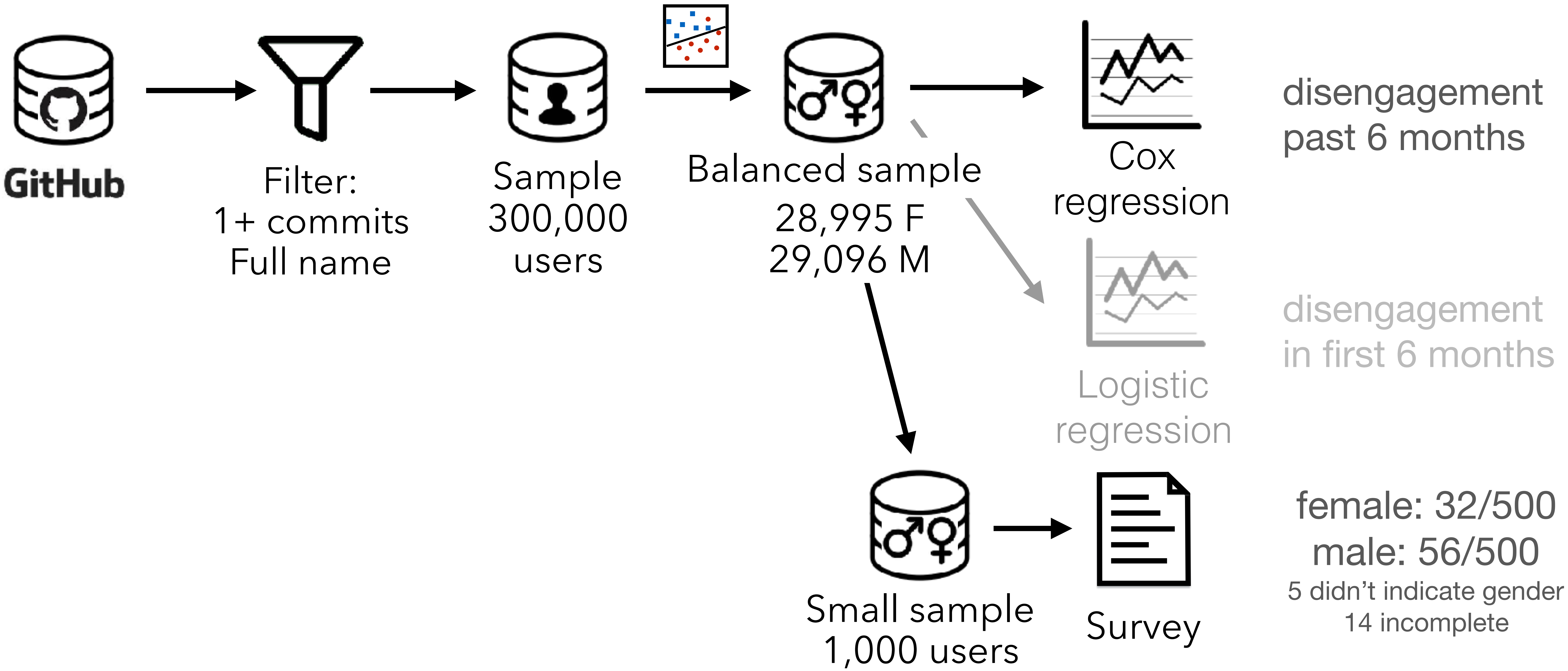
For the minority group, being attached to open teams helps to overcome the negative effects of network closure



Diversifying their ties makes women less dependent on the in-group for acceptance

Hypothesis: For women, higher chance of prolonged engagement with more diverse ties.

Large-scale mixed-methods study

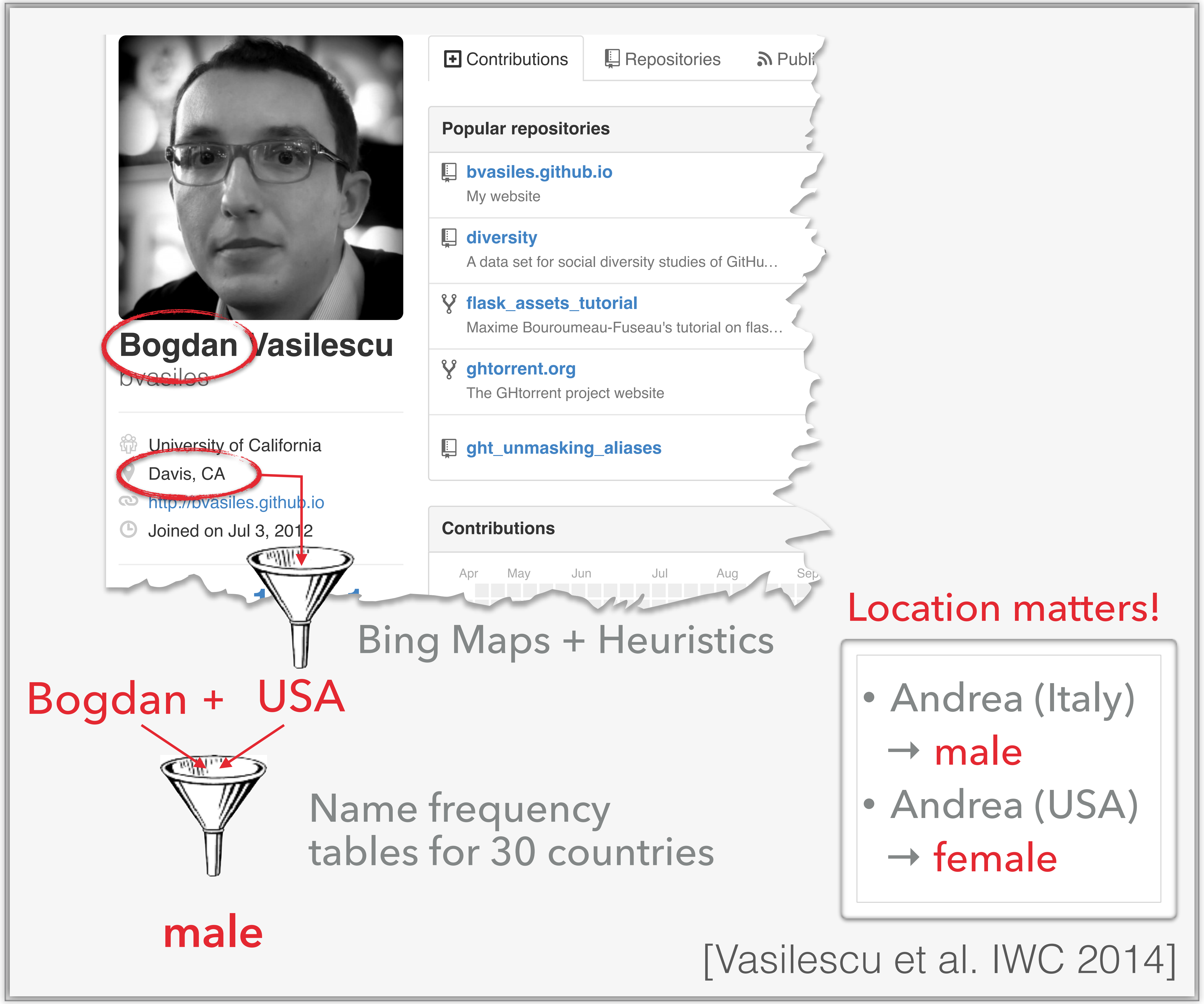


https://doi.org/10.5281/zenodo.2550931

Aside: Inferring gender from names

https://github.com/tue-mdse/genderComputer

gender Computer



Popular repositories

bvasiles.github.io
My website

diversity
A data set for social diversity studies of GitHu…

flask_assets_tutorial
Maxime Bouroumeau-Fuseau's tutorial on flas…

ghtorrent.org
The GHtorrent project website

ght_unmasking_aliases

Bogdan Vasilescu
bvasiles

University of California
Davis, CA
http://bvasiles.github.io
Joined on Jul 3, 2012

Contributions

Bing Maps + Heuristics

Bogdan + USA

Name frequency tables for 30 countries

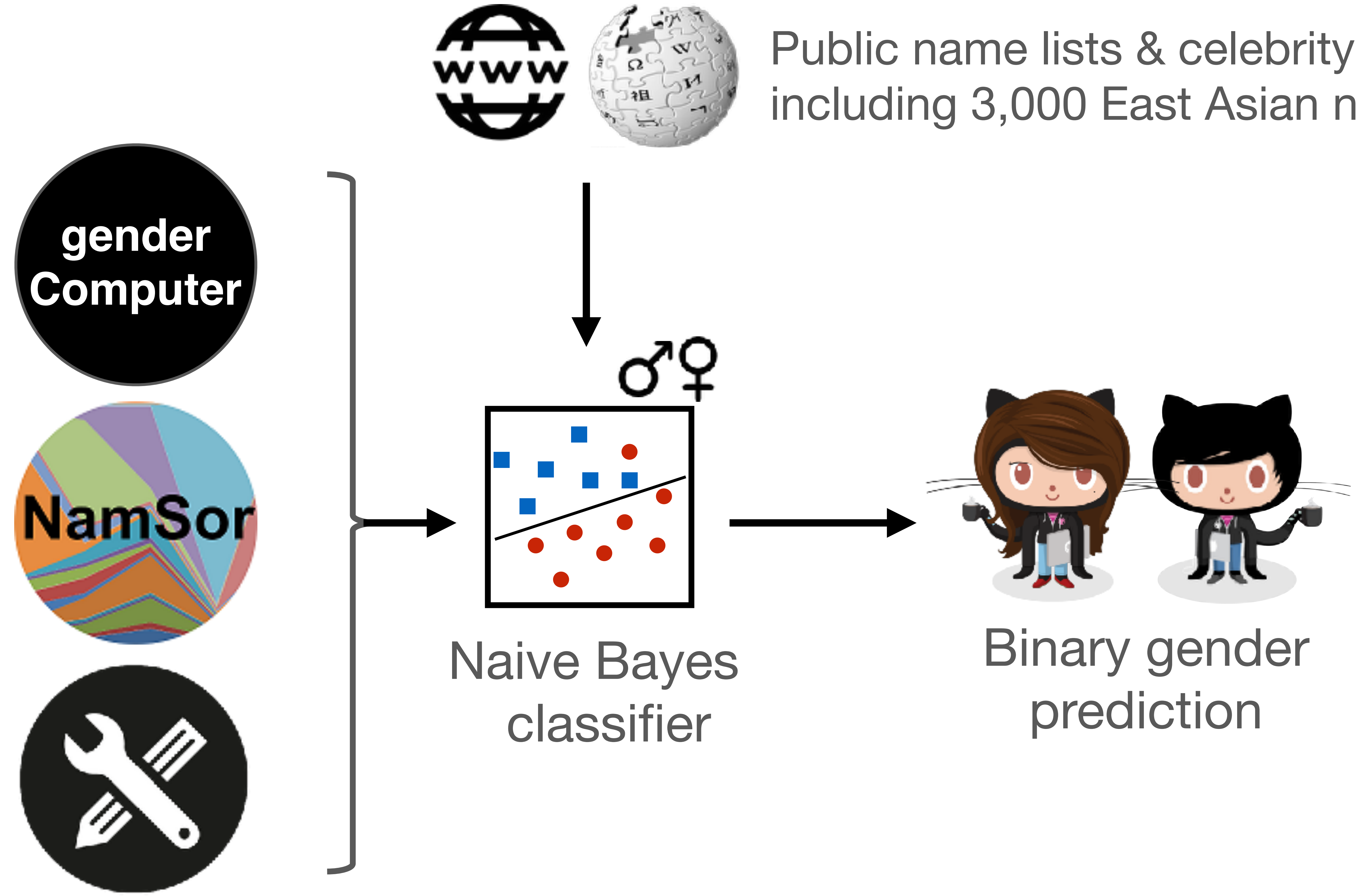
male

Location matters!

- Andrea (Italy) → male
- Andrea (USA) → female

[Vasilescu et al. IWC 2014]

Aside: Inferring gender from names



Public name lists & celebrity names, including 3,000 East Asian names

https://github.com/tue-mdse/genderComputer

https://www.namsor.com

name features, e.g., the last two characters

Naive Bayes classifier

Binary gender prediction

Aside: Inferring gender from names

	Accuracy		
Language	**genderComp.**	**NamSor**	**Our classifier**
Chinese	18%	7%	**60%**
Japanese	77%	27%	**80%**
Korean	19%	14%	**68%**
All	79%	74%	**84%**

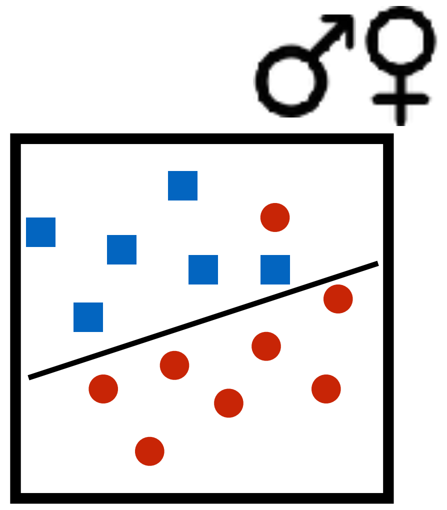
https://github.com/tue-mdse/genderComputer

gender Computer

https://www.namsor.com

NamSor

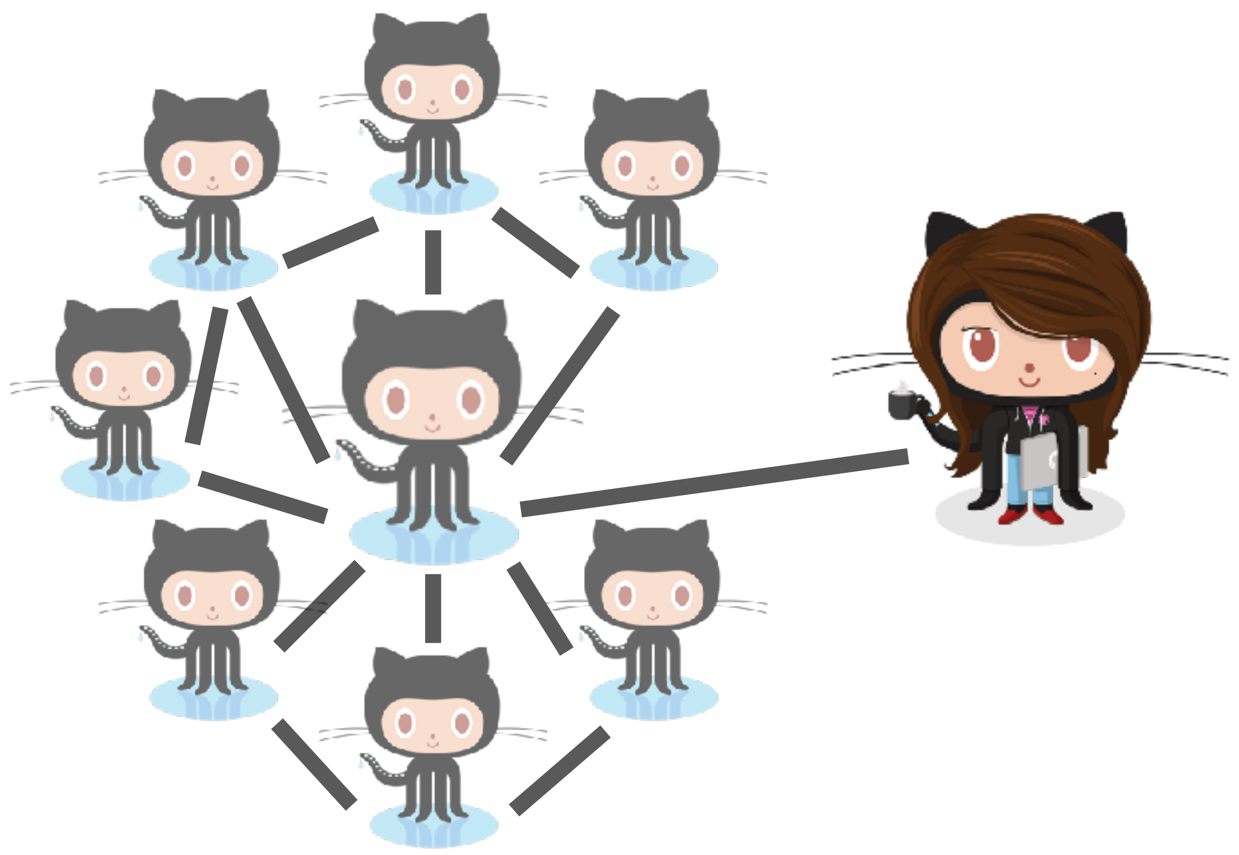
name features, e.g., the last two characters

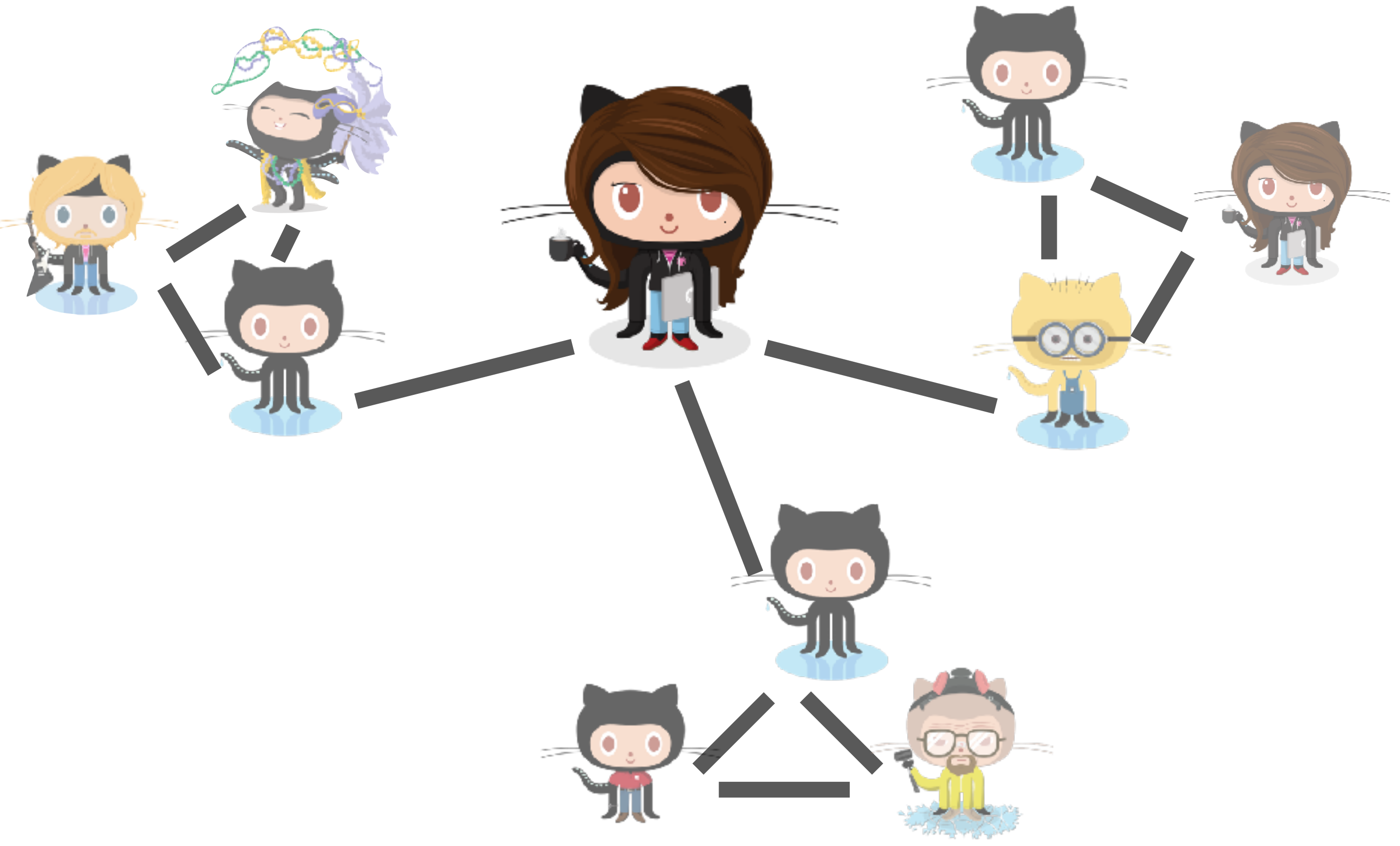


Naive Bayes classifier

Binary gender prediction

Operationalizations

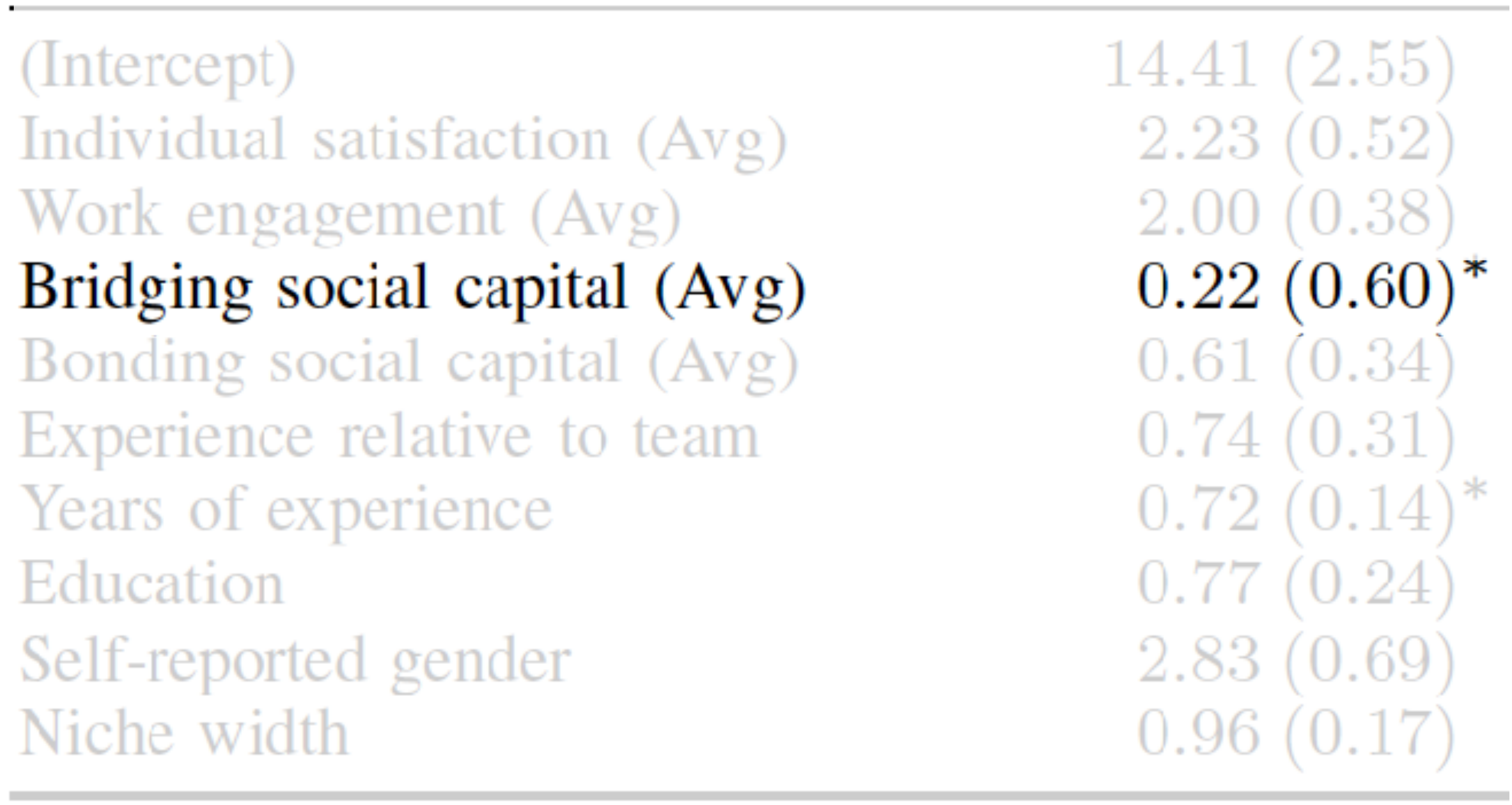




- *Disengagement*: no commits for 12 months
- Team cohesion (social capital)
  - *Team familiarity*: how well do you know people in a project on average, from previous projects (pairwise)
  - *Recurring cohesion*: cliques of at least three people who have previously worked together
- Information diversity of ties
  - *Share of newcomers*
  - *Heterogeneity of programming language expertise*: based on history of contributions to other projects
- Controls
  - Is project owner / major contributor (> 5% commits); followers; repository stars; niche width (programming languages)

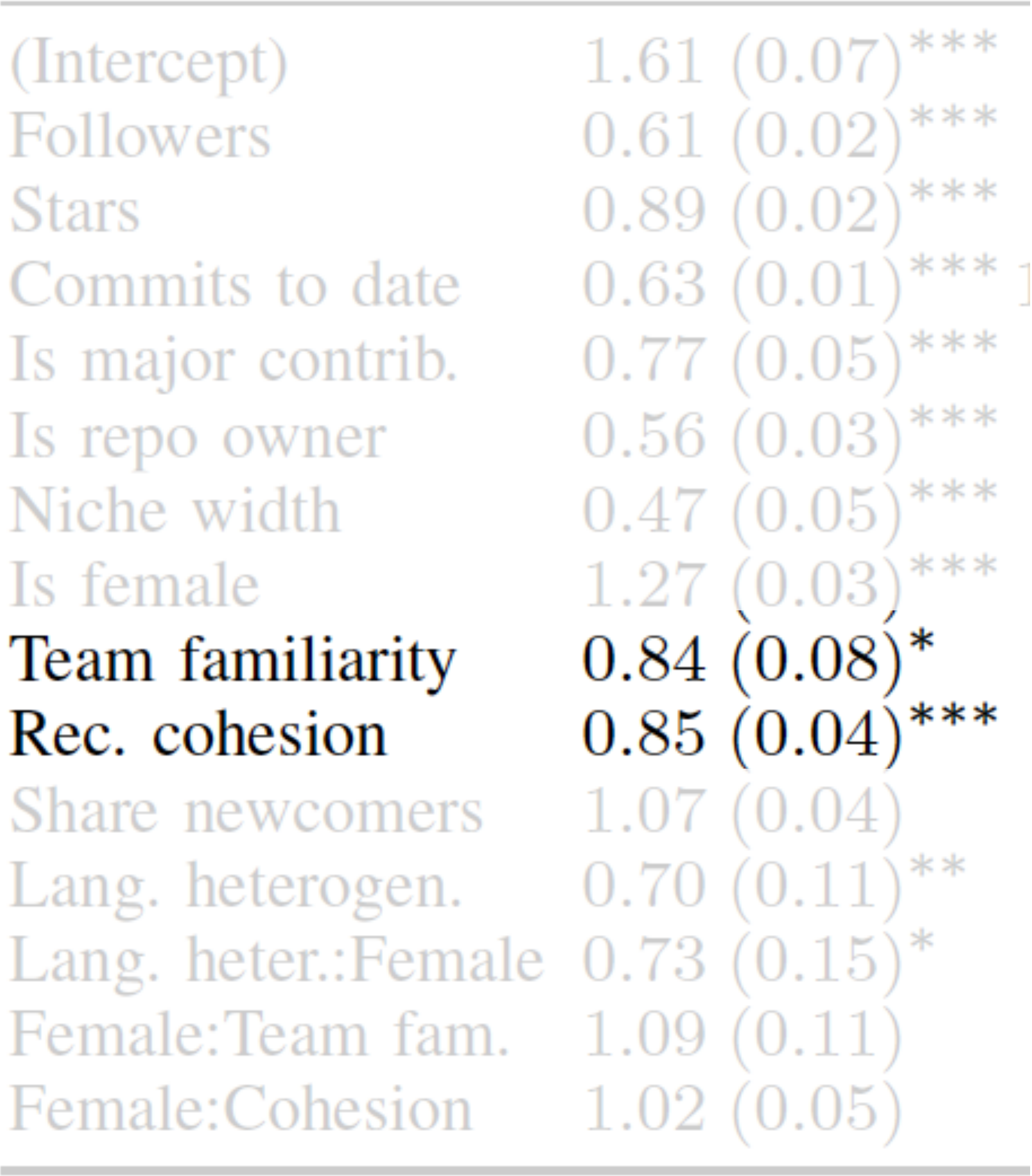
The more often people participate in projects with high potential for building social capital, the higher their chance of prolonged engagement

Survey



(Intercept)	14.41 (2.55)
Individual satisfaction (Avg)	2.23 (0.52)
Work engagement (Avg)	2.00 (0.38)
Bridging social capital (Avg)	**0.22 (0.60)***
Bonding social capital (Avg)	0.61 (0.34)
Experience relative to team	0.74 (0.31)
Years of experience	0.72 (0.14)*
Education	0.77 (0.24)
Self-reported gender	2.83 (0.69)
Niche width	0.96 (0.17)

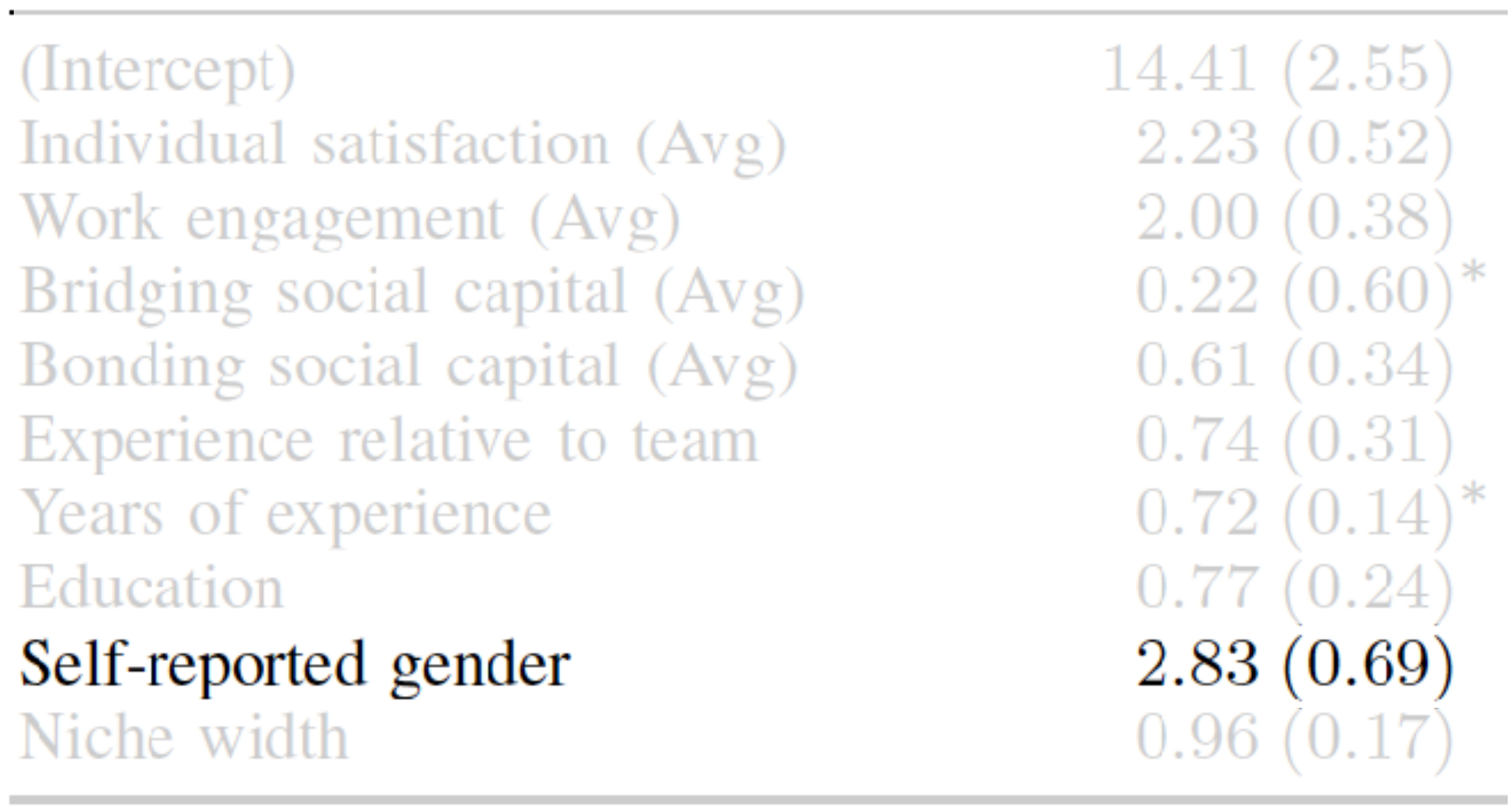
Repository mining



(Intercept)	1.61 (0.07)***
Followers	0.61 (0.02)***
Stars	0.89 (0.02)***
Commits to date	0.63 (0.01)***
Is major contrib.	0.77 (0.05)***
Is repo owner	0.56 (0.03)***
Niche width	0.47 (0.05)***
Is female	1.27 (0.03)***
Team familiarity	**0.84 (0.08)***
Rec. cohesion	**0.85 (0.04)*****
Share newcomers	1.07 (0.04)
Lang. heterogen.	0.70 (0.11)**
Lang. heter.:Female	0.73 (0.15)*
Female:Team fam.	1.09 (0.11)
Female:Cohesion	1.02 (0.05)

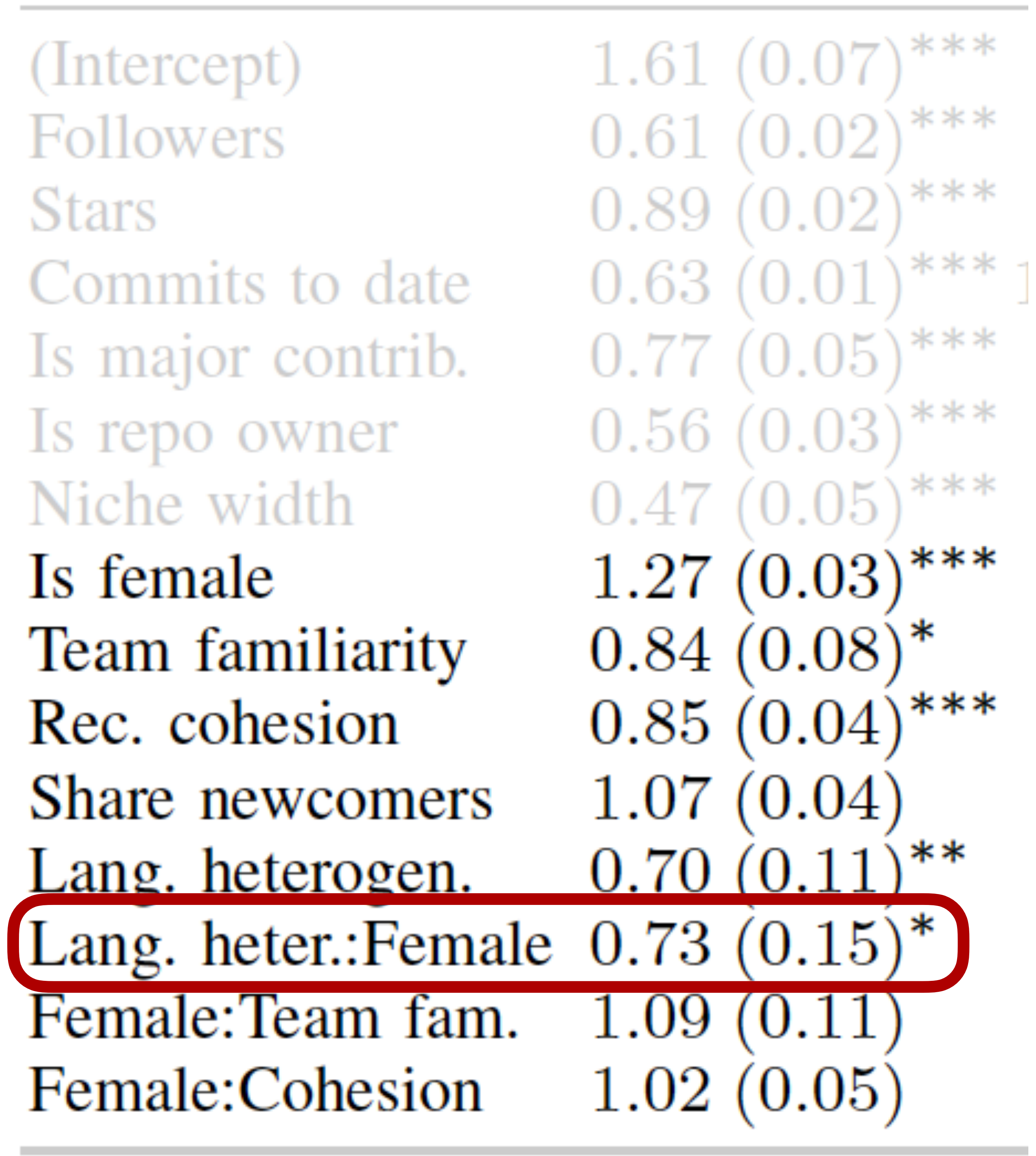
Language heterogeneity interacts with gender

Survey



(Intercept)	14.41 (2.55)
Individual satisfaction (Avg)	2.23 (0.52)
Work engagement (Avg)	2.00 (0.38)
Bridging social capital (Avg)	0.22 (0.60)*
Bonding social capital (Avg)	0.61 (0.34)
Experience relative to team	0.74 (0.31)
Years of experience	0.72 (0.14)*
Education	0.77 (0.24)
Self-reported gender	2.83 (0.69)
Niche width	0.96 (0.17)

Repository mining



(Intercept)	1.61 (0.07)***
Followers	0.61 (0.02)***
Stars	0.89 (0.02)***
Commits to date	0.63 (0.01)***
Is major contrib.	0.77 (0.05)***
Is repo owner	0.56 (0.03)***
Niche width	0.47 (0.05)***
Is female	1.27 (0.03)***
Team familiarity	0.84 (0.08)*
Rec. cohesion	0.85 (0.04)***
Share newcomers	1.07 (0.04)
Lang. heterogen.	0.70 (0.11)**
Lang. heter.:Female	0.73 (0.15)*
Female:Team fam.	1.09 (0.11)
Female:Cohesion	1.02 (0.05)

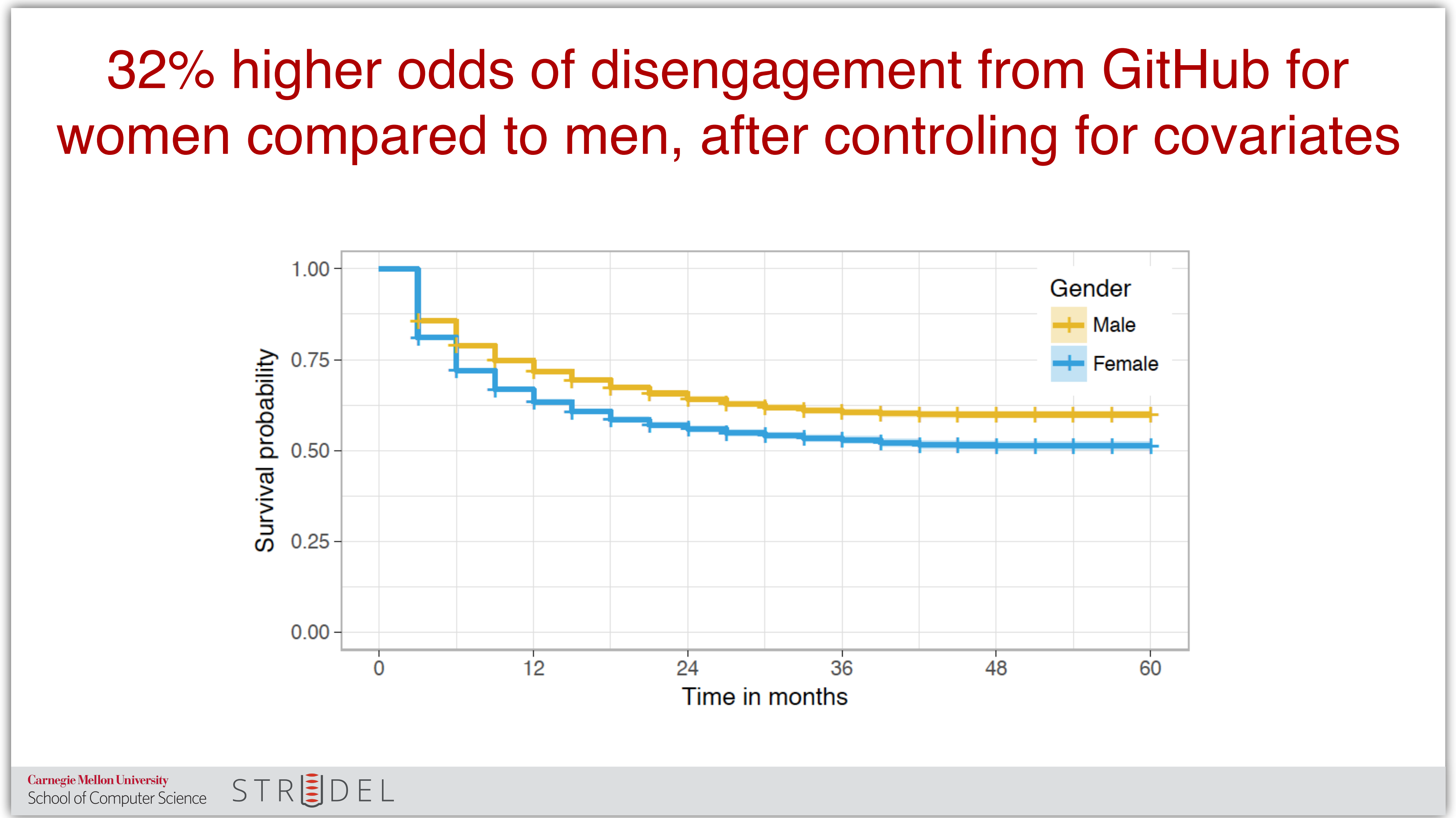
Women are more likely to disengage when language heterogeneity is low

Women disengage for personal reasons significantly more often than men

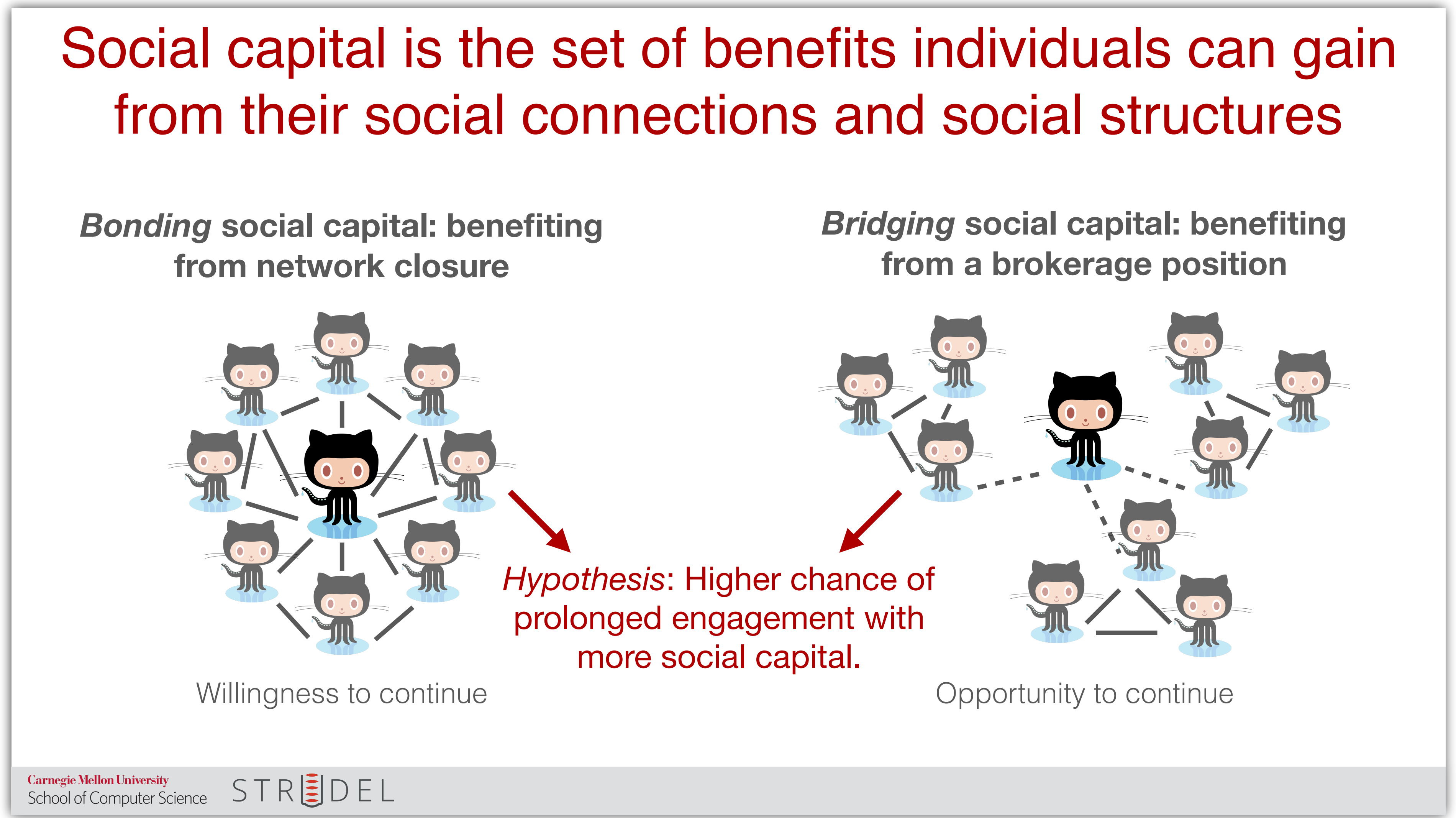
- Common self-reported reasons for disengaging:
  - lack of time
    - work related (“changes in job”, “work became overbearing”)
    - personal reasons (“diversifying hobbies”, “personal life”)
  - no personal need for that software anymore

Survey

Social capital theory is a useful framework to study contributor (dis)engagement in open source



32% higher odds of disengagement from GitHub for women compared to men, after controling for covariates



Social capital is the set of benefits individuals can gain from their social connections and social structures

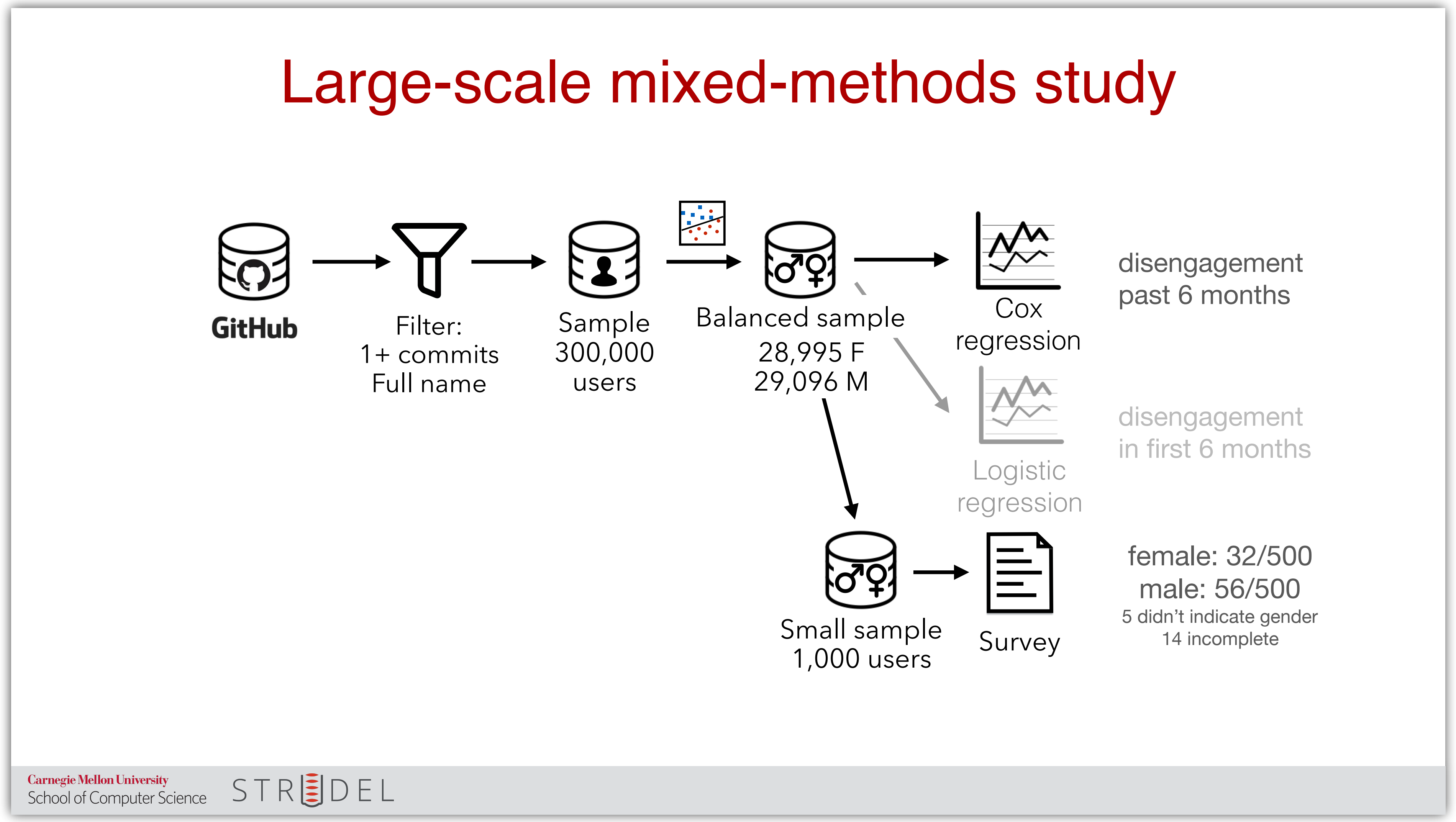
Bonding **social capital: benefiting from network closure**

Bridging **social capital: benefiting from a brokerage position**

Hypothesis: Higher chance of prolonged engagement with more social capital.

Willingness to continue

Opportunity to continue

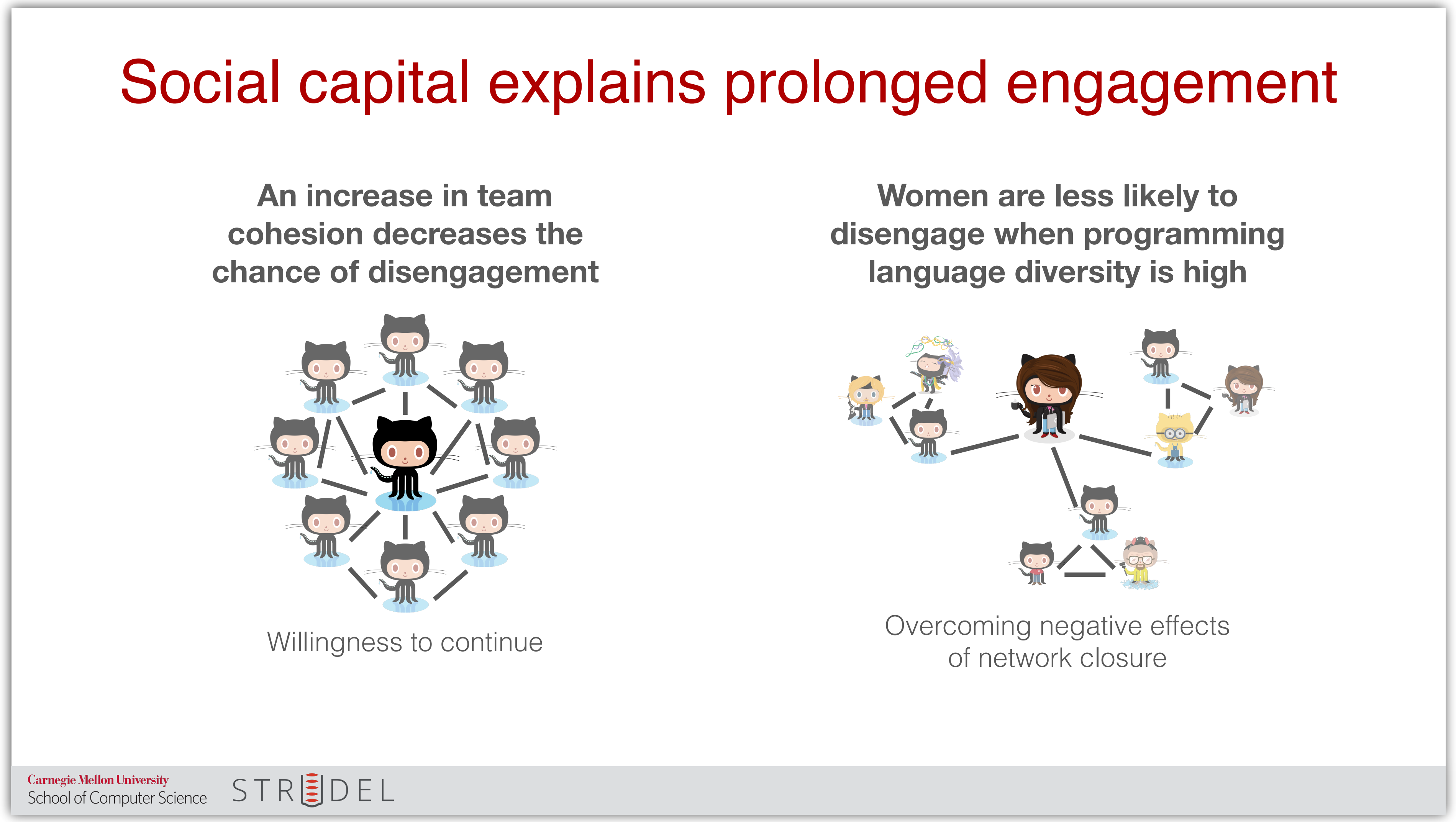


Large-scale mixed-methods study

disengagement past 6 months

disengagement in first 6 months

female: 32/500
male: 56/500
5 didn't indicate gender
14 incomplete



Social capital explains prolonged engagement

An increase in team cohesion decreases the chance of disengagement

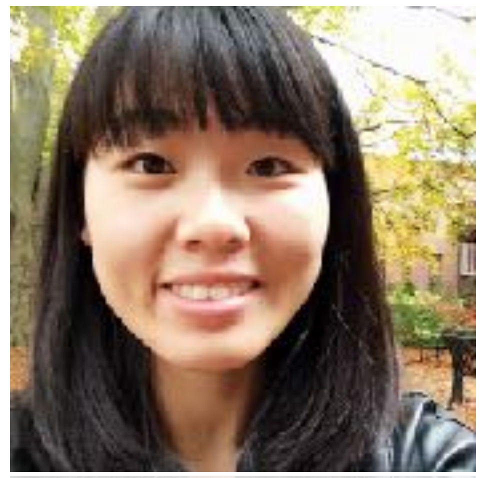
Women are less likely to disengage when programming language diversity is high

Willingness to continue

Overcoming negative effects of network closure

Acknowledgements

Anita Brown



Michelle Cao

Jim Herbsleb

Christian Kästner

Alex Nolte

Sophie Qiu

Alex Serebrenik

Marat Valiev

Open source needs a steady supply of time and effort by contributors

But that is harder today than ever before

… because of how open source has changed

Many more questions we need answers to

- Which open source projects form digital infrastructure
- How open source digital infrastructure is being used
- How much and what kind of effort does each project need
- **How do project interdependencies impact sustainability**
- How do people choose which projects to contribute to
- How to attract a more diverse pool of contributors
- **Why do open source contributors disengage** / how to retain them
- Which project-level practices and policies encourage contributions
- How effective are the different support models / what are their side effects
- How much can transparency help the ecosystem to self regulate