DALL·E 3 - "A selection of strudel pastries with different fillings, including apple, sour cherry, and poppy seed"

Sustainable Open-Source Ecosystems:

What We've Learned So Far and the Road Ahead

Bogdan Vasilescu

SIESTA Summer School, September 4, 2024

STRUDEL
SOCIO-TECHNICAL RESEARCH USING DATA EXCAVATION LAB

Carnegie Mellon University

S3D
Software and Societal Systems Department

About me

@b_vasilescu

Associate Professor @CMU

Director of the Societal Computing PhD program

STRUDEL research group

Societal Computing

Software and Societal Systems Department

Integrate and Innovate

Empower change as a tech-driven leader, shaping a more equitable world through our interdisciplinary PhD program in Societal Computing.

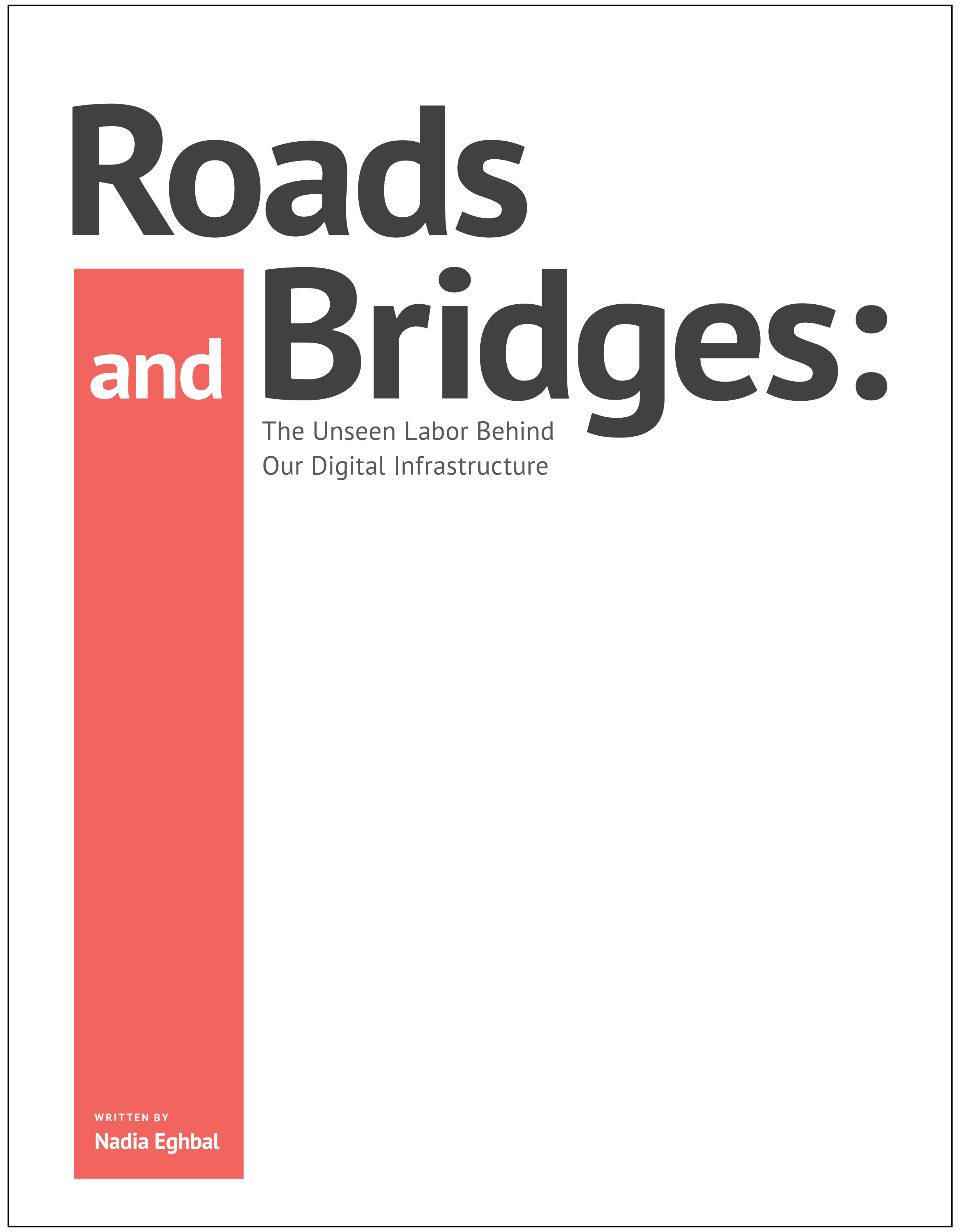
Shaping the Digital Future

Welcome to the Societal Computing Program at Carnegie Mellon University, where we tackle the complex challenges at the intersection of computation, society, and policy. With a multi-disciplinary approach, our PhD program prepares tomorrow's leaders to design technologies addressing societal needs and guide their implementation. Explore our cutting-edge research, innovative curriculum, and join us in shaping the

Our Program

Our PhD program in the integrated, innovative discipline of Societal Computing provides the techniques, theories, and research methods to address societal issues and create technologies that impact society.

Open source software has become digital infrastructure



Roads and Bridges:

The Unseen Labor Behind Our Digital Infrastructure

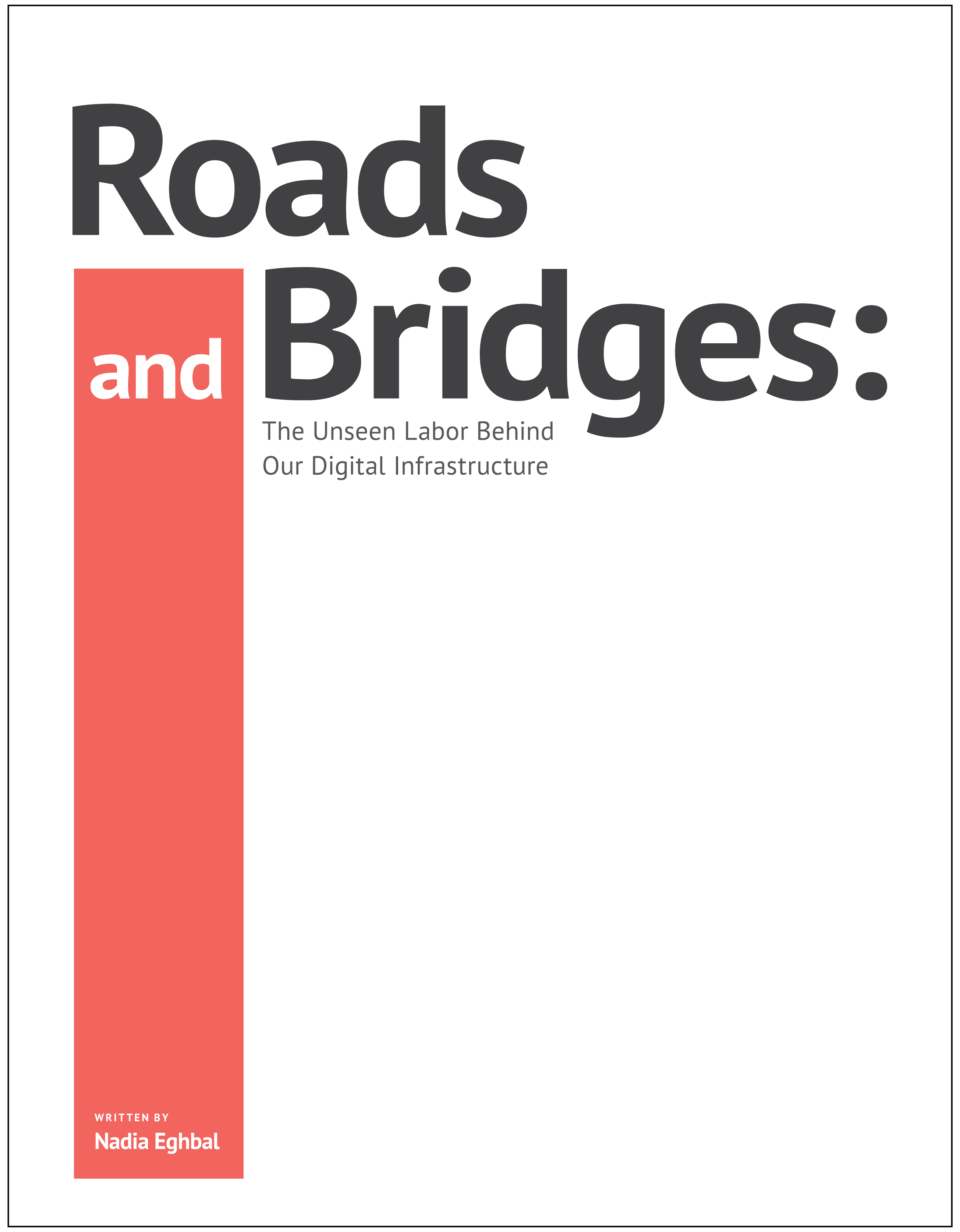
WRITTEN BY

Nadia Eghbal

Everybody uses open source:

- Fortune 500 companies
- Major software companies
- Startups
- Government
- …

Like any infrastructure, it needs regular upkeep and maintenance



Roads and Bridges:

The Unseen Labor Behind Our Digital Infrastructure

WRITTEN BY

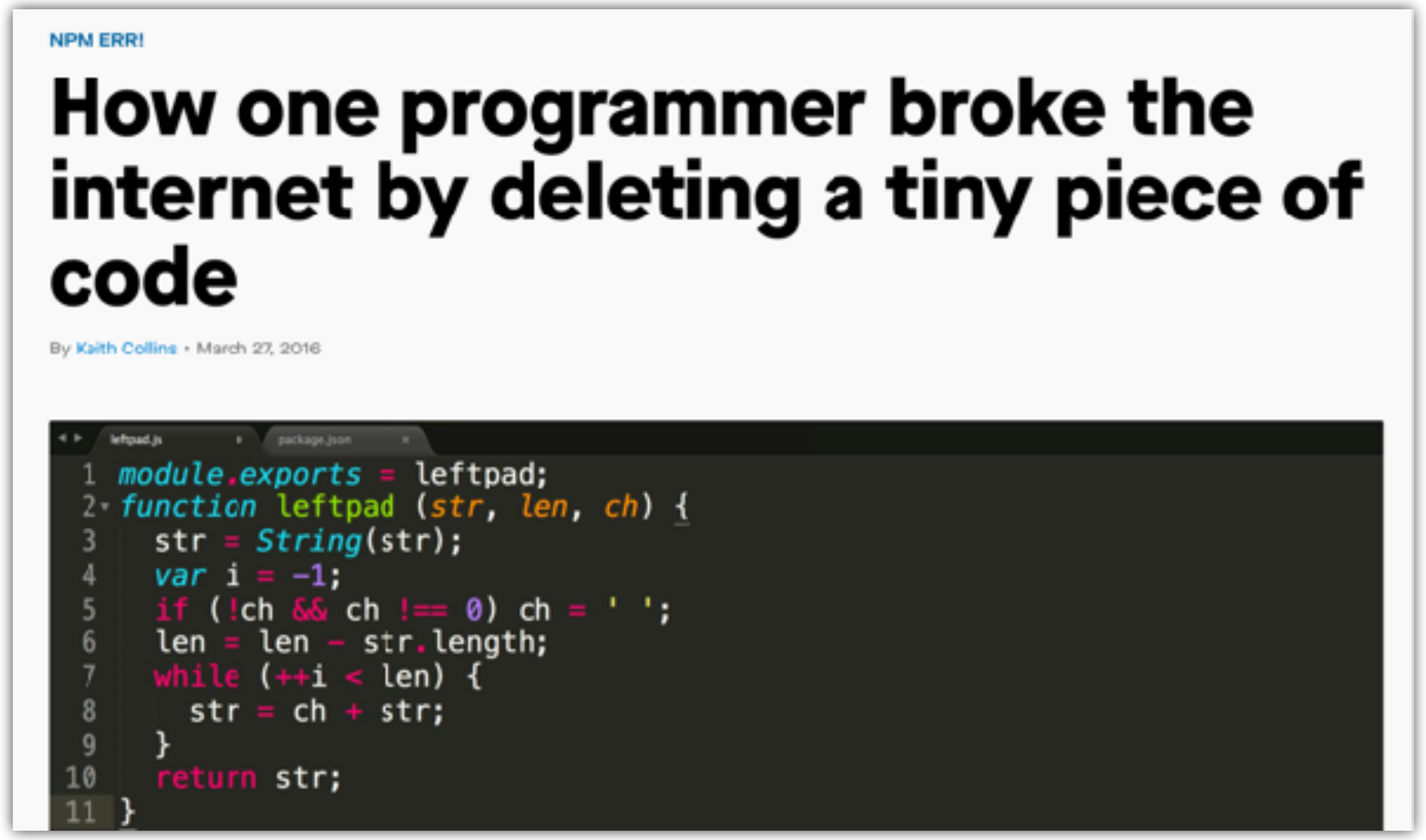
Nadia Eghbal

Everybody uses open source:

- Fortune 500 companies
- Major software companies
- Startups
- Government
- …

If undermaintained:

- Brittle supply chains
- Risks for downstream users
- Slows down innovation
- …



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/

Sustaining open source is hard

Ever more open source software is being created (and reused)

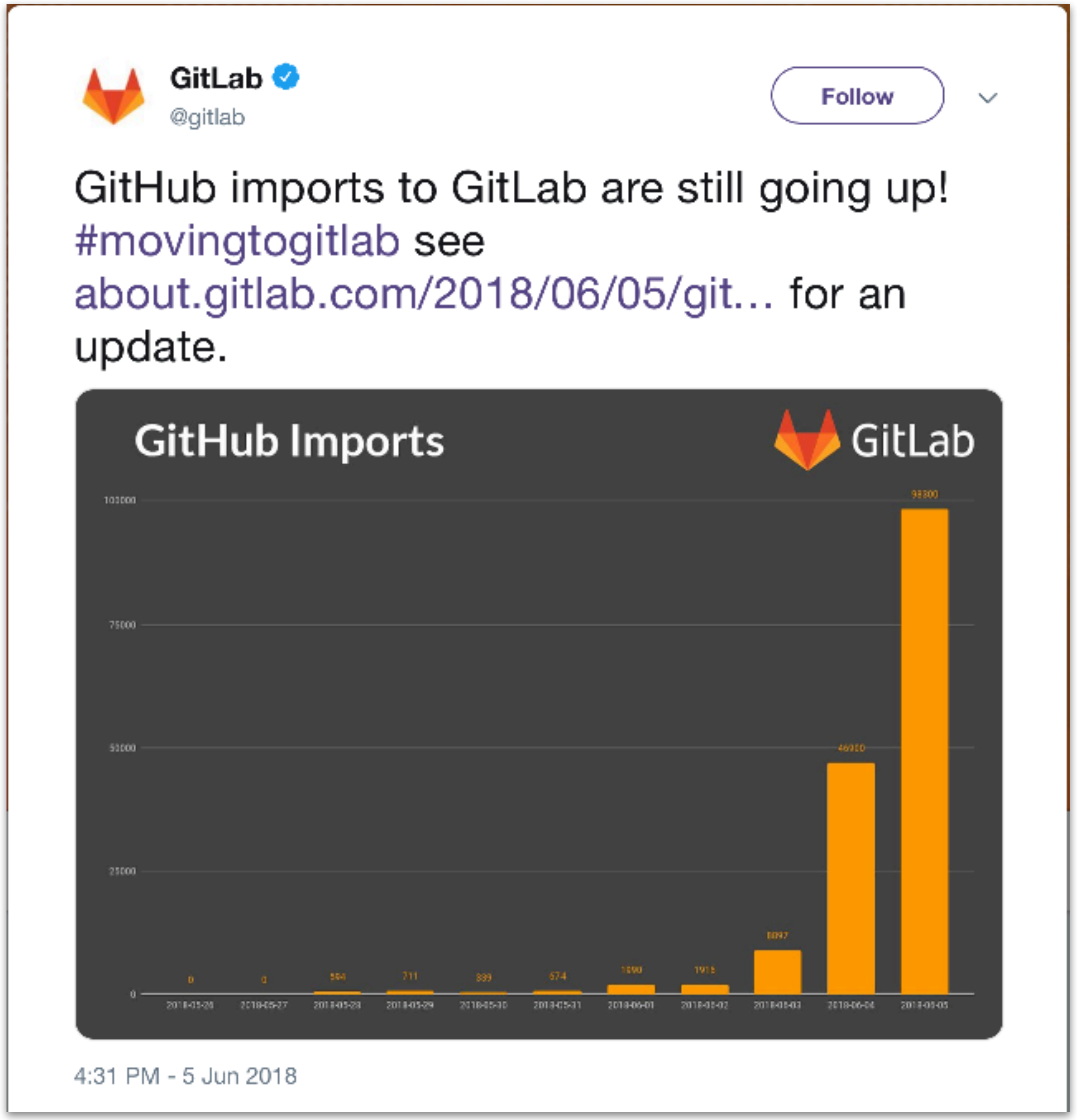
Explosion of production in the past 10 years

github SOCIAL CODING

400+ million repositories
100+ million users
(February 2024)

Bitbucket

10+ million users
(April 2019)



GitLab
@gitlab

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

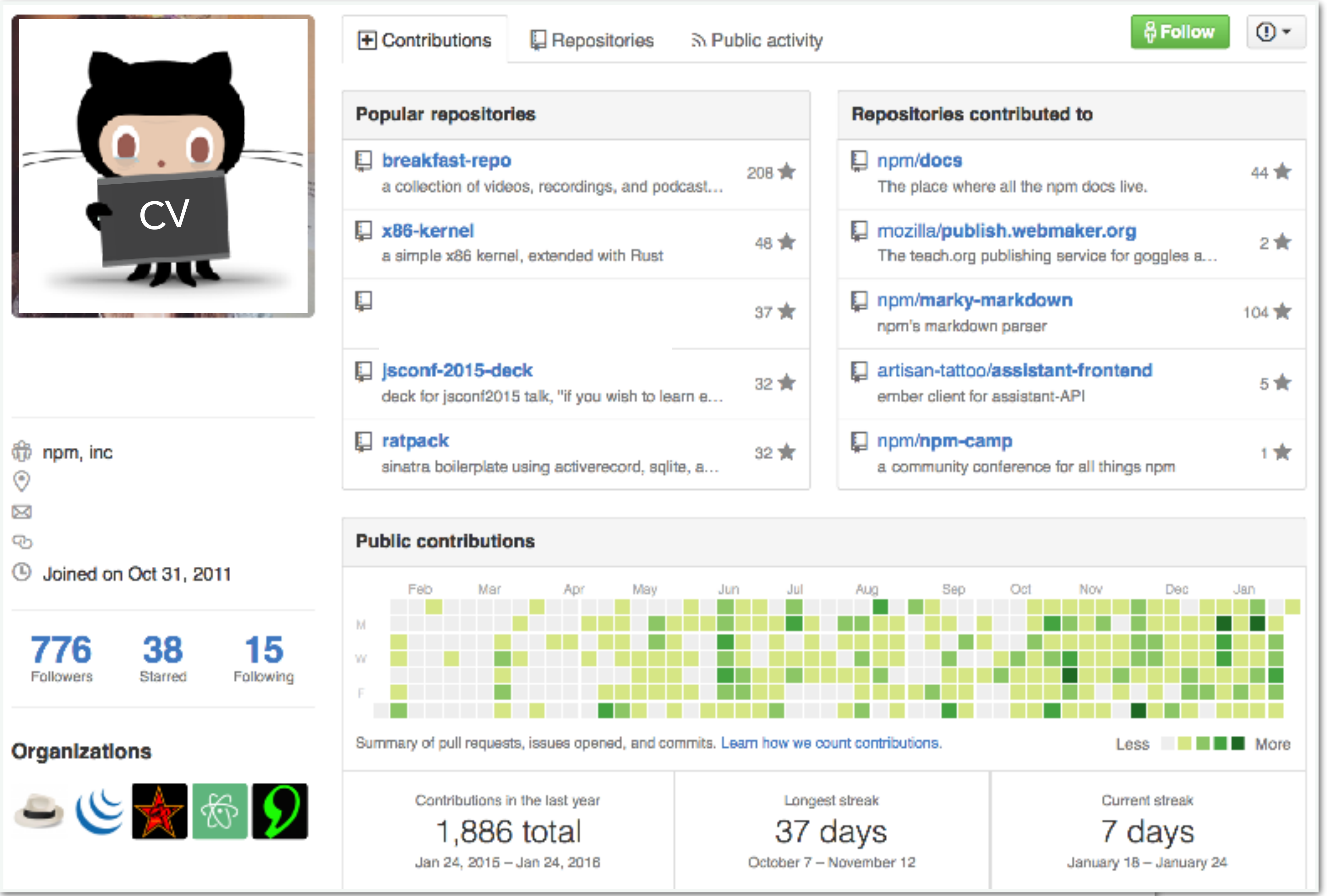
The social platforms have won

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)



Popular repositories

breakfast-repo 208 ★
a collection of videos, recordings, and podcast...

x86-kernel 48 ★
a simple x86 kernel, extended with Rust

37 ★

jsconf-2015-deck 32 ★
deck for jsconf2015 talk, "if you wish to learn e...

ratpack 32 ★
sinatra boilerplate using activerecord, sqlite, a...

Repositories contributed to

npm/docs 44 ★
The place where all the npm docs live.

mozilla/publish.webmaker.org 2 ★
The teach.org publishing service for goggles a...

npm/marky-markdown 104 ★
npm's markdown parser

artisan-tattoo/assistant-frontend 5 ★
ember client for assistant-API

npm/npm-camp 1 ★
a community conference for all things npm

npm, inc

Joined on Oct 31, 2011

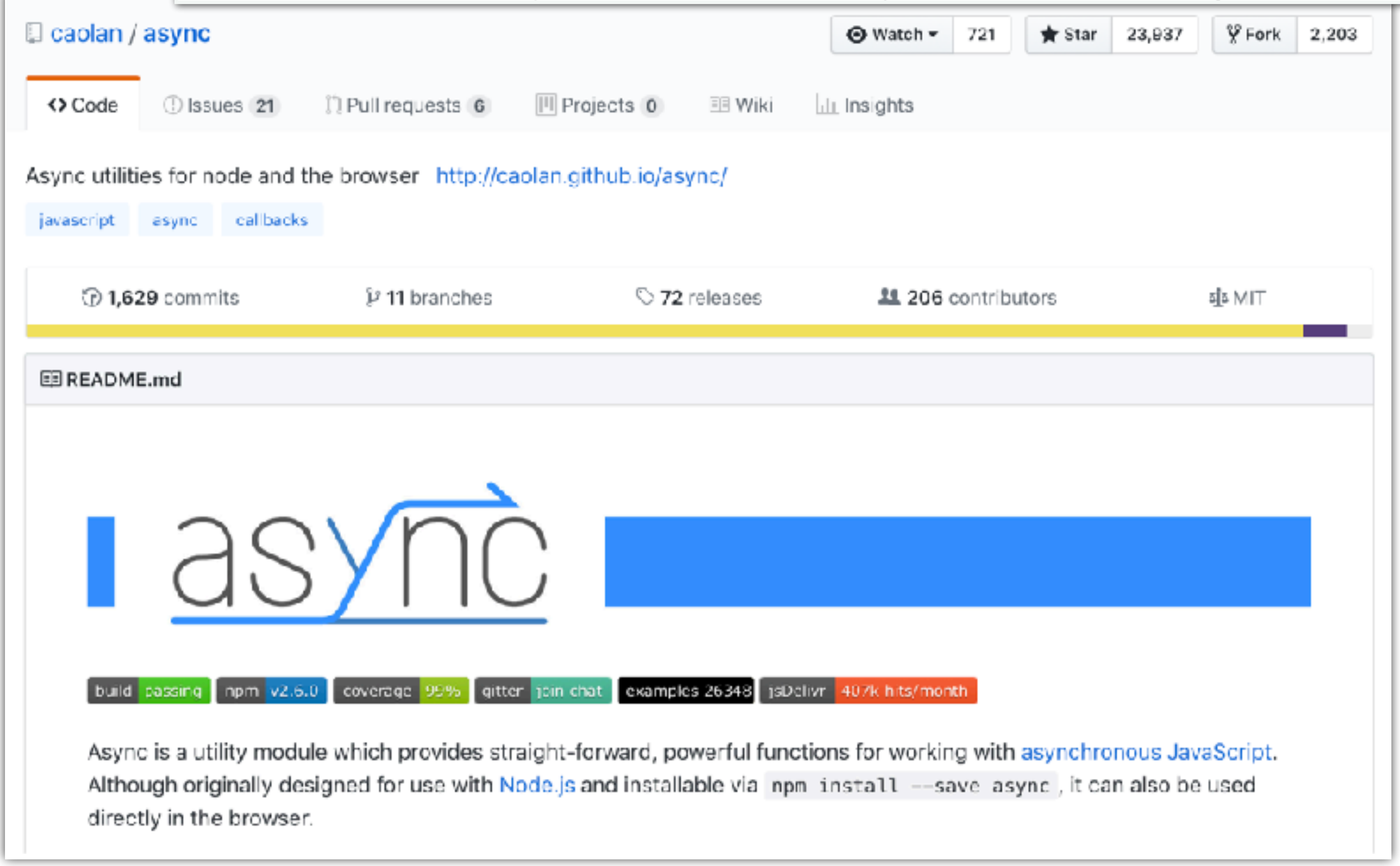
776 Followers · 38 Starred · 15 Following

Organizations

Public contributions

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

Contributions in the last year	Longest streak	Current streak
1,886 total	37 days	7 days
Jan 24, 2015 – Jan 24, 2016	October 7 – November 12	January 18 – January 24



caolan / async

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 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.

There is increasing commercialization and professionalization

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

- More recently, 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/

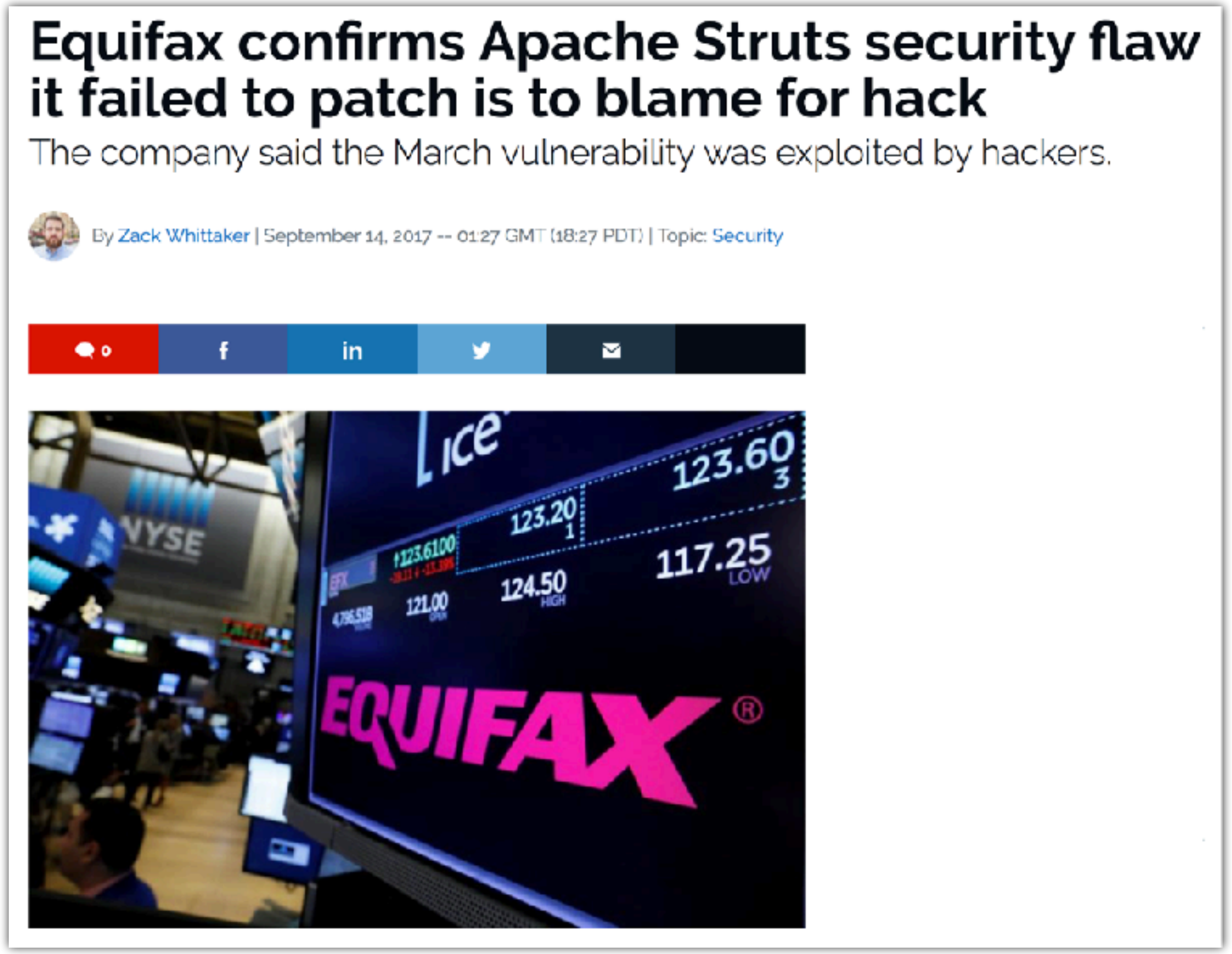
Expectations toward the quality, reliability, and security of open source infrastructure are high

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

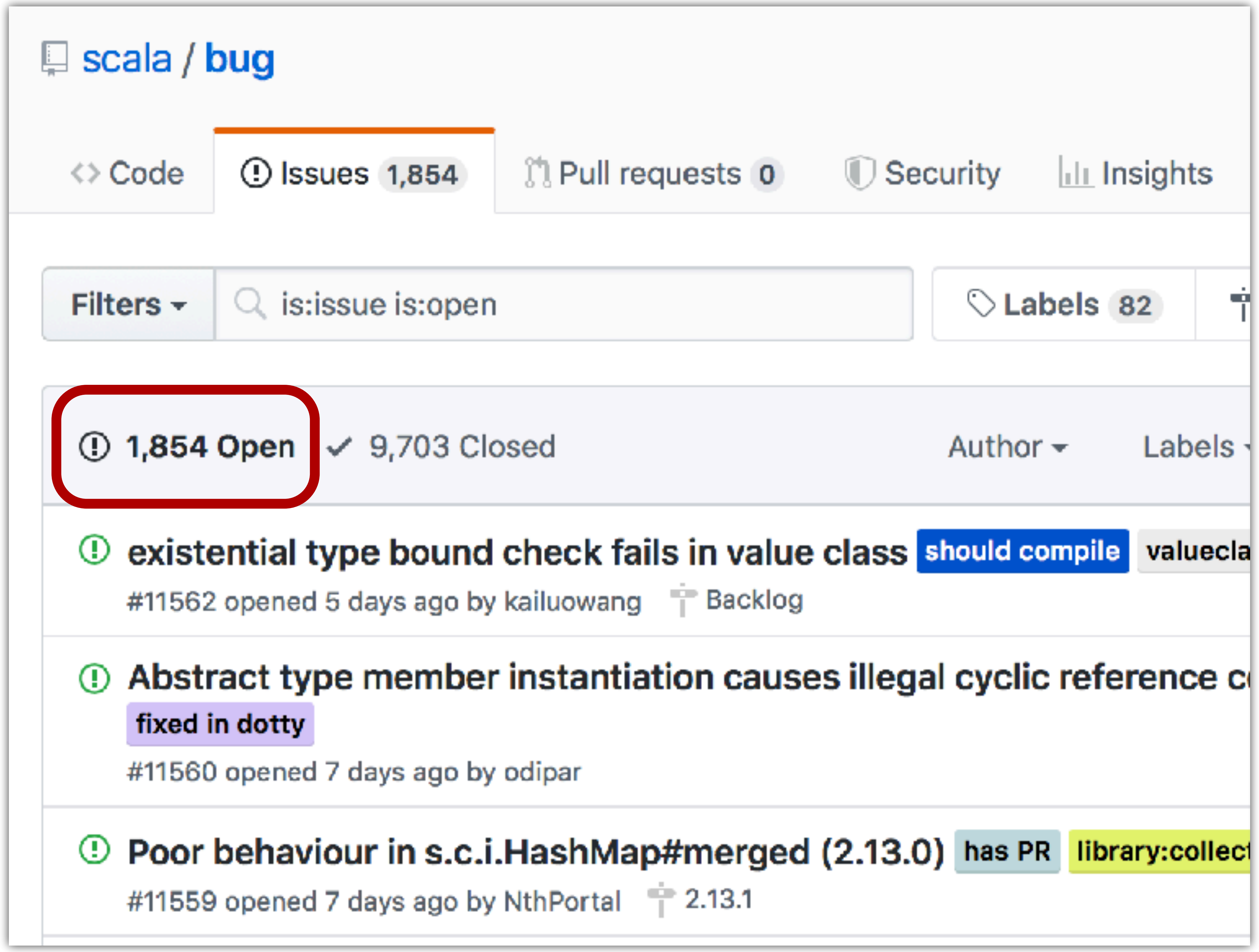
- Growing volume of requests

Social pressure to respond quickly

- Otherwise, off-putting to newcomers (Steinmacher et al. 2015)

Entitled, unreasonable 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.”*



is:issue is:open

1,854 Open ✓ 9,703 Closed

existential type bound check fails in value class should compile valuecla
#11562 opened 5 days ago by kailuowang Backlog

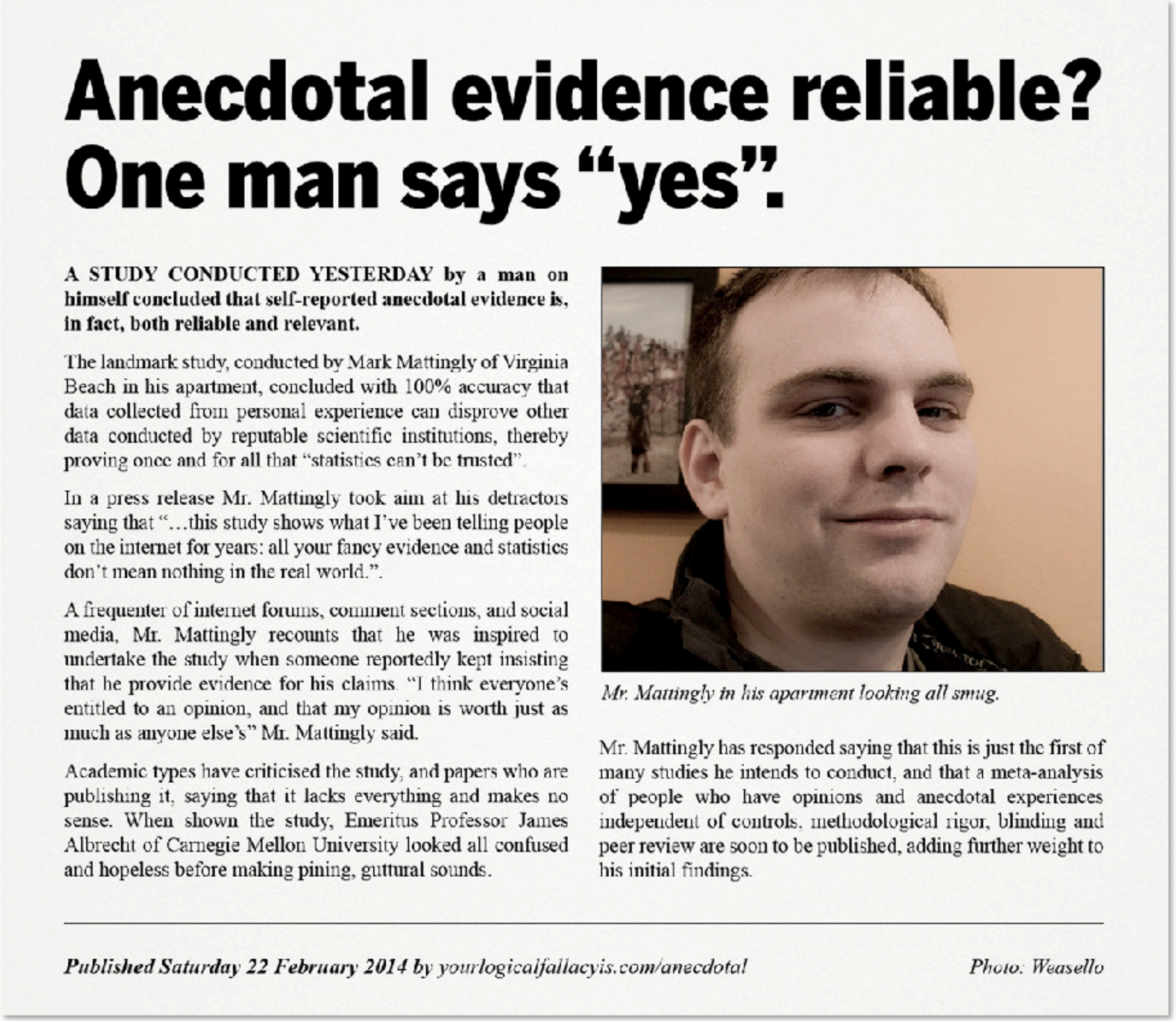
Abstract type member instantiation causes illegal cyclic reference c
fixed in dotty
#11560 opened 7 days ago by odipar

Poor behaviour in s.c.i.HashMap#merged (2.13.0) has PR library:collect
#11559 opened 7 days ago by NthPortal 2.13.1

Science is needed for evidence-based recommendations

Lots of change
Lots of challenges

Little evidence or theory



Anecdotal evidence reliable? One man says "yes".

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

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

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

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

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

Mr. Mattingly in his apartment looking all smug.

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

Published Saturday 22 February 2014 by yourlogicalfallacyis.com/anecdotal

Photo: Weasello

A great opportunity for research!

A great opportunity for research!

… because (almost) everything being archived and public makes it possible to study the problem empirically

A great opportunity for research!

… because (almost) everything being archived and public makes it possible to study the problem empirically

“The collection of public Git repositories as a whole […] exceeds 1.5PB” (Ma et al, 2021)

Ma, Y., Dey, T., Bogart, C., Amreen, S., Valiev, M., Tutko, A., ... & Mockus, A. (2021). World of code: enabling a research workflow for mining and analyzing the universe of open source VCS data. Empirical Software Engineering, 26, 1-42.

Today: Let’s look at three concrete examples

Dealing with abandoned upstream dependencies

Estimating a project’s effective labor pool

Estimating causal effects of promotional activities

A Closer Look at Abandonment

C. Miller, C. Kästner, and B. Vasilescu. “We feel like we’re winging it:” A study on navigating open-source dependency abandonment. In International Conference on the Foundations of Software Engineering (FSE), page 1281–1293. ACM, 2023.

C. Miller, M. Jahanshahi, A. Mockus, B. Vasilescu, and C. Kästner. Understanding the Response to Open-Source Dependency Abandonment in the npm Ecosystem. In International Conference on Software Engineering (ICSE). IEEE, 2025.

DALL·E 3 - An abandoned bakery. Rotten strudel pastries are lying around

Most prior research has focused on keeping projects “alive” and maintained.

- Attracting and onboarding new contributors
- Reducing barriers to entry
- Improving the culture
- Improving funding models

…

Maintainers often leave projects for reasons we can't / shouldn’t prevent

- Switching jobs (voluntarily)
- Starting a family
- Losing interest

…

Research should also focus on helping open-source maintainers with sunsetting, and helping open-source users with the effects of that.

How big is the problem?
What do people do to prepare / deal with it?

Interviews with maintainers of Javascript, Python, and PHP projects with abandoned upstream dependencies.

A large-scale quantitative study of abandoned npm packages.

DALL·E 3 - An abandoned bakery. Rotten strudel pastries are lying around

https://github.com/dimsemenov/Magnific-Popup
Additions and Deletions per week
25K
20K
15K
10K
5K
0
5K
10K
15K
January 2014
January 2015
January 2016
January 2017
January 2018
January 2019
January 2020
January 2021
January 2022
January 2023
January 2024

https://github.com/dimsemenov/Magnific-Popup

Additions and Deletions per week
25K
20K
15K
10K
5K
0
5K
10K
15K
January 2014
January 2015
January 2016
January 2017
January 2018
January 2019
January 2020
January 2021
January 2022
January 2023
January 2024

Considered “abandoned” here

2+ years of activity

2+ years of complete inactivity

https://github.com/dimsemenov/Magnific-Popup
Additions and Deletions per week
25K
20K
15K
10K
5K
0
5K
10K
15K
January 2014
January 2015
January 2016
January 2017
January 2018
January 2019
January 2020
January 2021
January 2022
January 2023
January 2024

```
+ **Important note** This jQuery plugin is deprecated, only critical or
  security bug fixes will be released in future. Use native `<dialog>`
  element if you need a basic dialog/modal/popup, or my <a
  href="https://photoswipe.com">PhotoSwipe</a> library if you need an
  advanced image gallery. Feel free to email me if you need assistance.
```

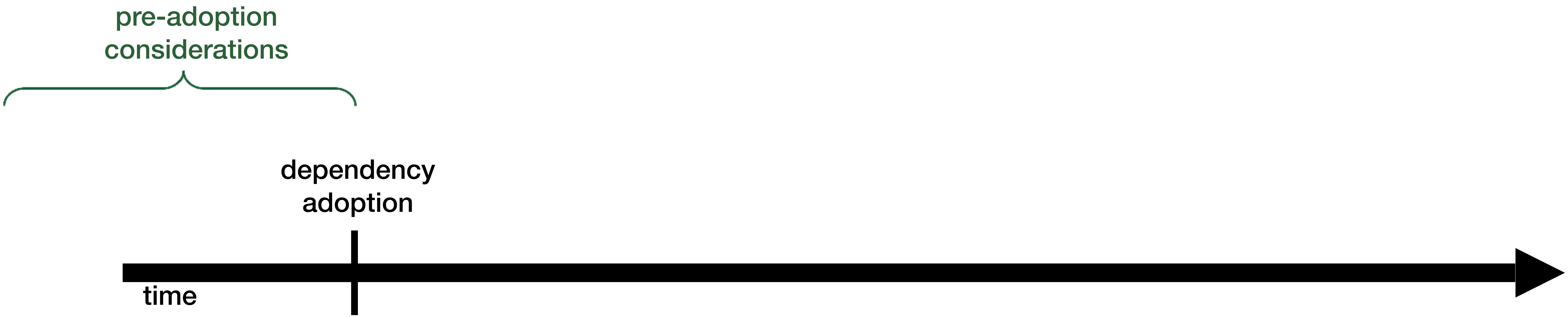
Considered “abandoned” here

2+ years of activity

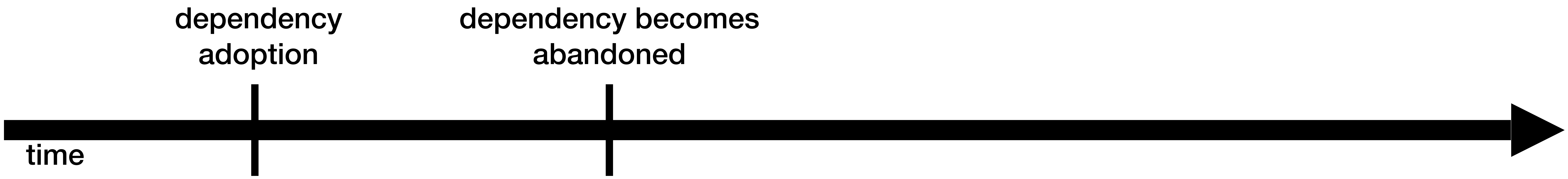
2+ years of complete inactivity

Part 1: Interviews

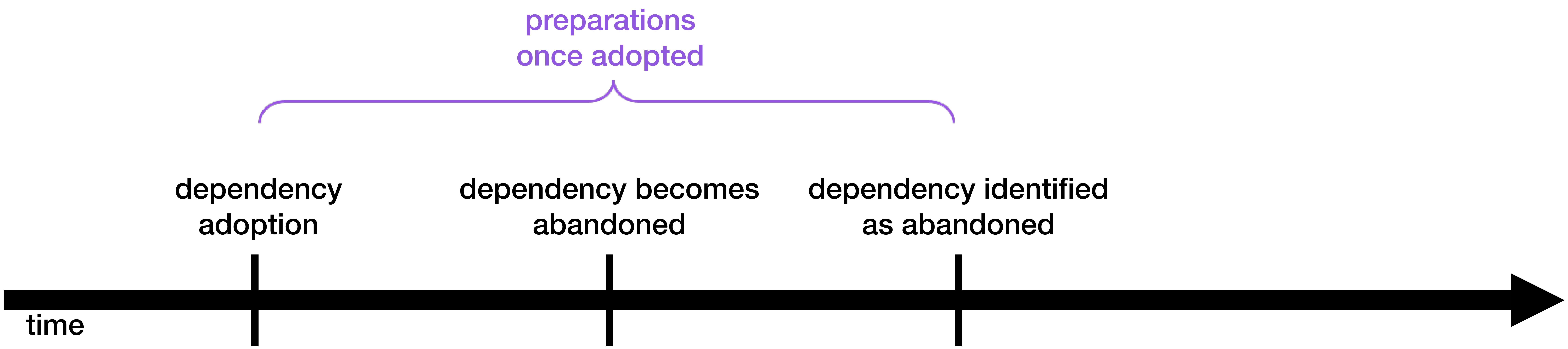
Timeline from the perspective of a consumer



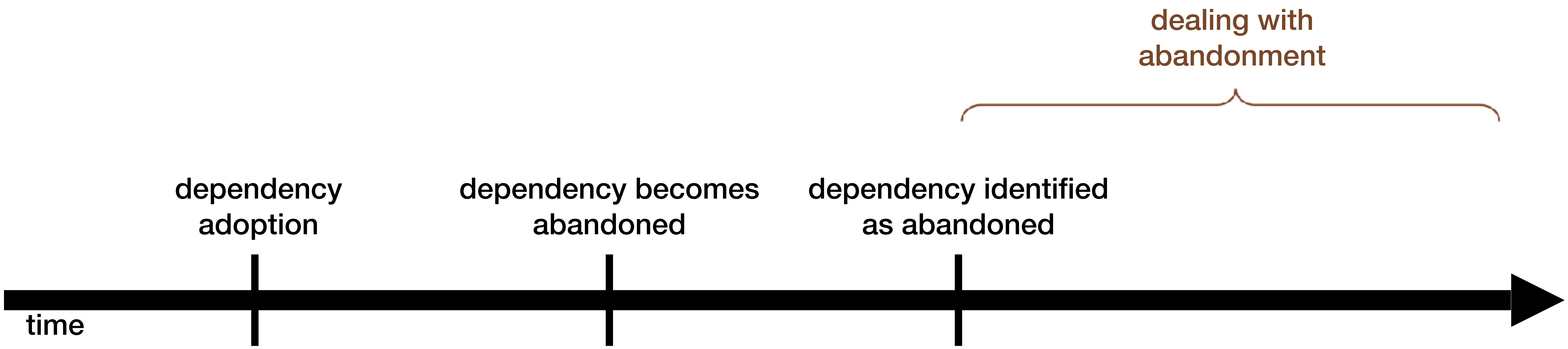
Timeline from the perspective of a consumer



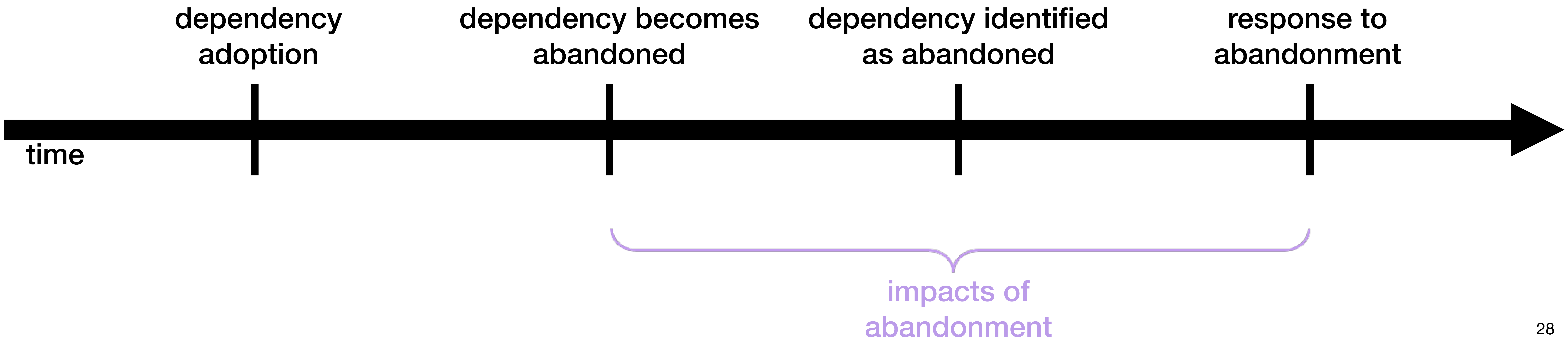
Timeline from the perspective of a consumer



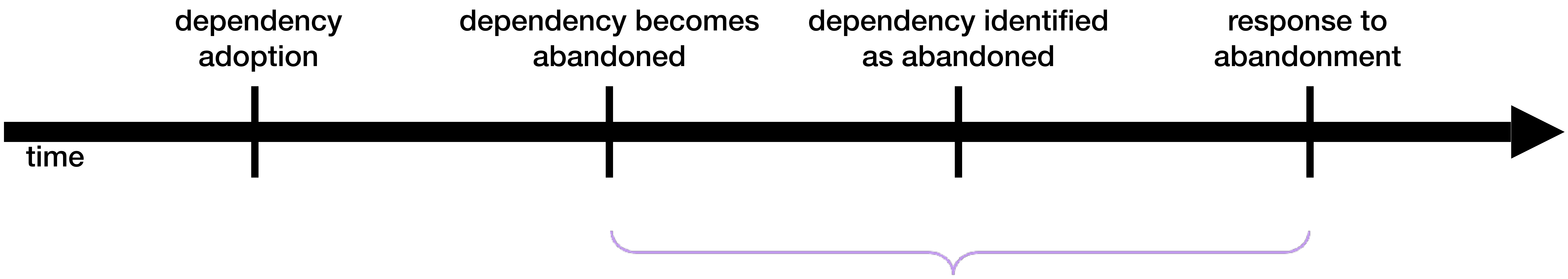
Timeline from the perspective of a consumer


dealing with abandonment
dependency adoption
dependency becomes abandoned
dependency identified as abandoned
time

Timeline from the perspective of a consumer



Impacts of abandonment are debated



dependency adoption

dependency becomes abandoned

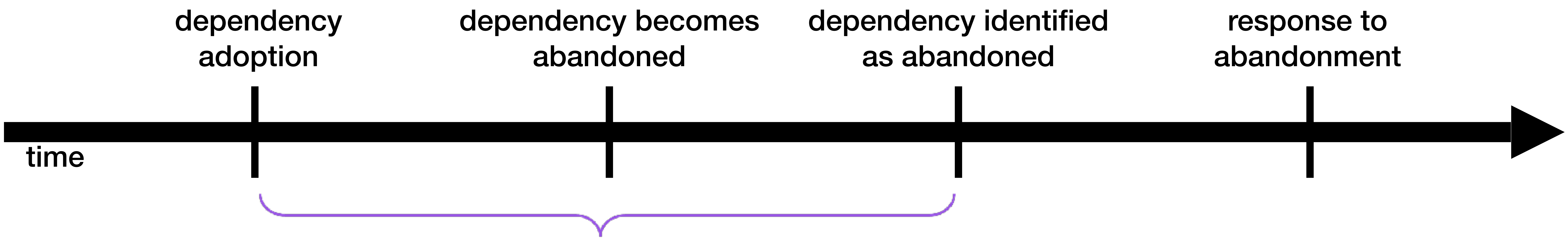
dependency identified as abandoned

response to abandonment

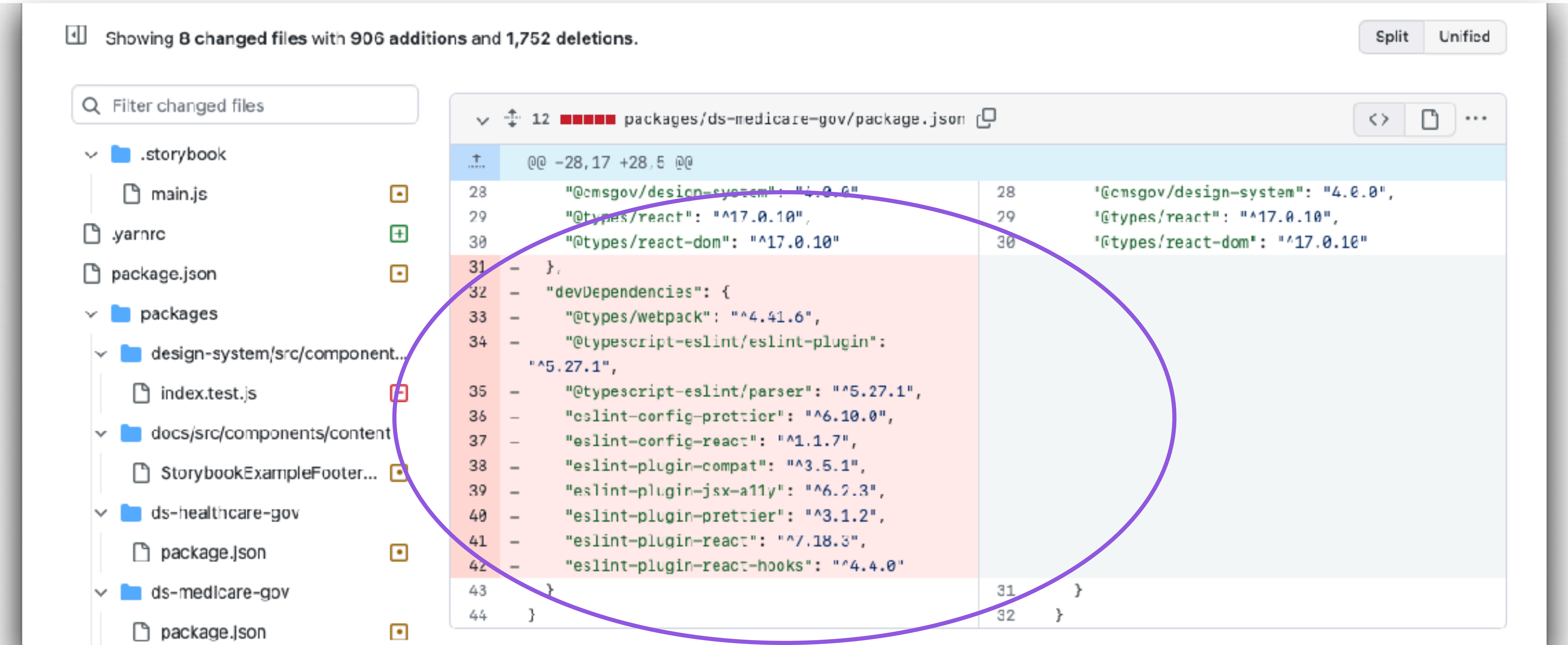
time

- Some concrete, e.g., language incompatibilities (Python 2 to 3), missing needed features
- Many more anticipated, e.g., future updates, security concerns
- Some expect no meaningful impact

Preparations post-adoption seem rare


dependency adoption
dependency becomes abandoned
dependency identified as abandoned
response to abandonment
time

E.g., building abstraction layers, minimizing dependencies, monitoring



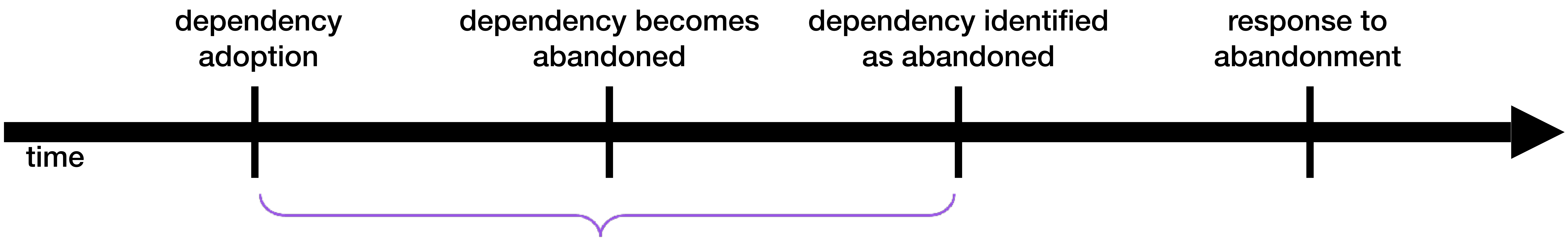
Showing 8 changed files with 906 additions and 1,752 deletions.

- .storybook
  - main.js
- .yarnrc
- package.json
- packages
  - design-system/src/component...
    - index.test.js
  - docs/src/components/content
    - StorybookExampleFooter...
  - ds-healthcare-gov
    - package.json
  - ds-medicare-gov
    - package.json

12 packages/ds-medicare-gov/package.json

```
@@ -28,17 +28,5 @@
       "@cmsgov/design-system": "4.0.0",
       "@types/react": "^17.0.10",
       "@types/react-dom": "^17.0.10"
-    },
-    "devDependencies": {
-      "@types/webpack": "^4.41.6",
-      "@typescript-eslint/eslint-plugin":
   "^5.27.1",
-      "@typescript-eslint/parser": "^5.27.1",
-      "eslint-config-prettier": "^6.10.0",
-      "eslint-config-react": "^1.1.7",
-      "eslint-plugin-compat": "^3.5.1",
-      "eslint-plugin-jsx-a11y": "^6.2.3",
-      "eslint-plugin-prettier": "^3.1.2",
-      "eslint-plugin-react": "^7.18.3",
-      "eslint-plugin-react-hooks": "^4.4.0"
     }
   }
```

Preparations post-adoption seem rare



dependency adoption

dependency becomes abandoned

dependency identified as abandoned

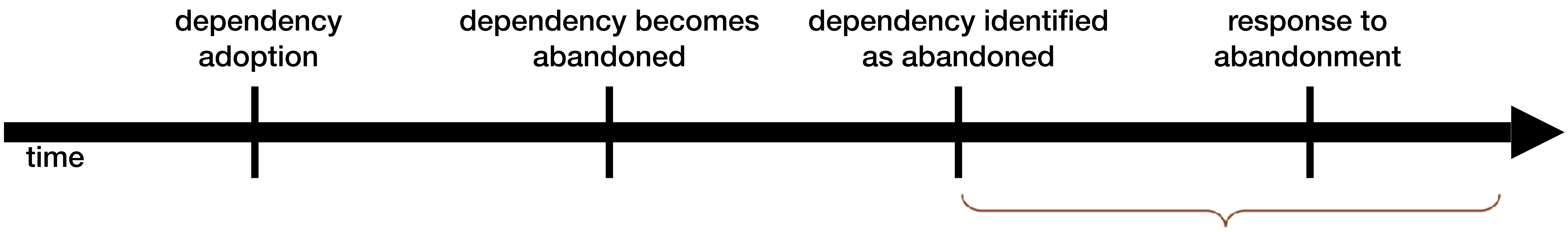
response to abandonment

time

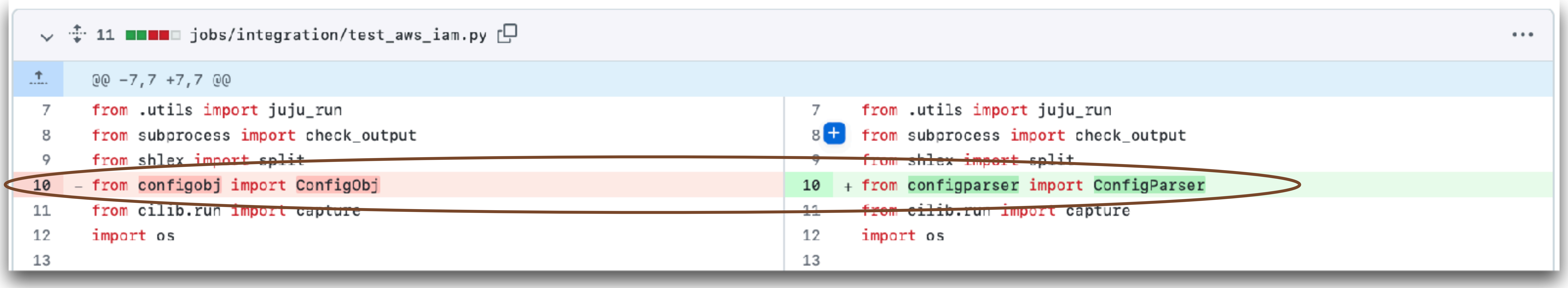
Not all interviewees considered prep worth the effort

> *We are basically employing the strategy of*
>
> ***'if it works it works, if it breaks then I'll fix the issues.'***
>
> *- PID10*

The most common way to deal with abandonment is to switch to an alternative dependency


dependency adoption
dependency becomes abandoned
dependency identified as abandoned
response to abandonment
time

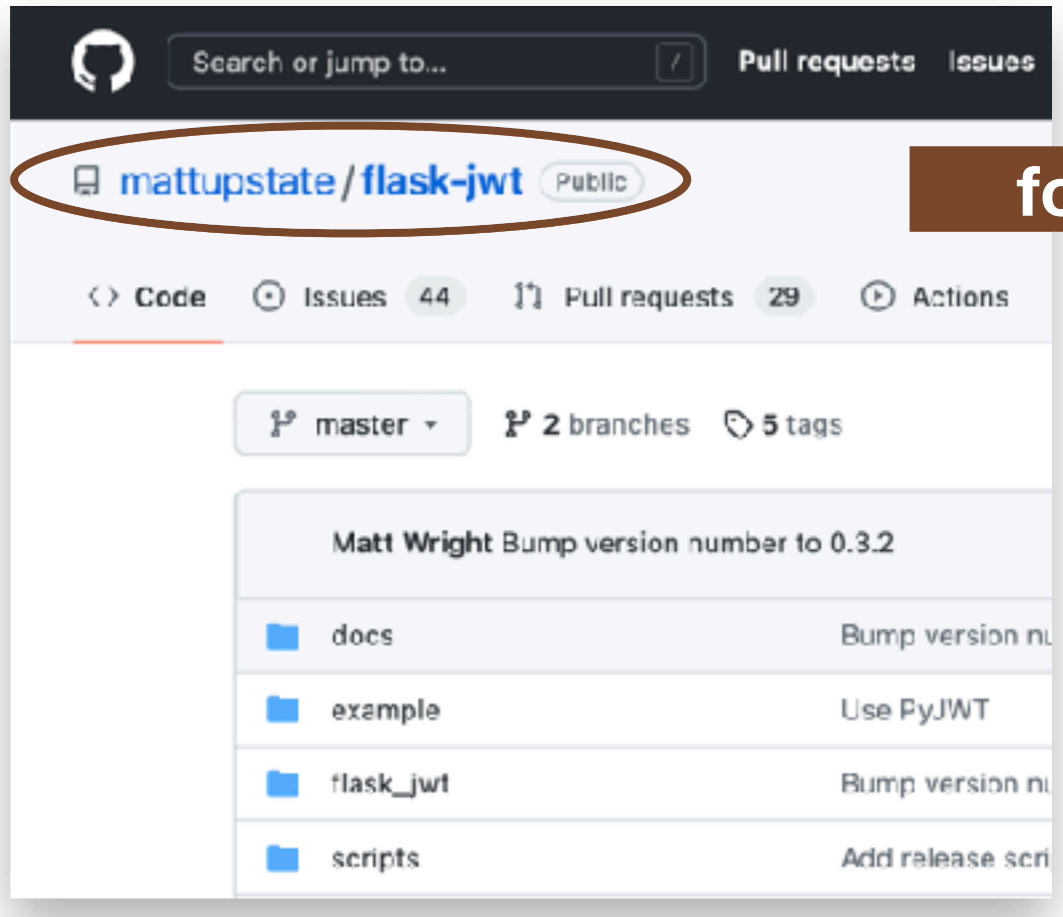
Another common solution was to fork or vendor code



```
11 jobs/integration/test_aws_iam.py
@@ -7,7 +7,7 @@
 from .utils import juju_run
 from subprocess import check_output
 from shlex import split
-from configobj import ConfigObj
+from configparser import ConfigParser
 from cilib.run import capture
 import os

```

Dealing with abandonment typically required trial-and-error

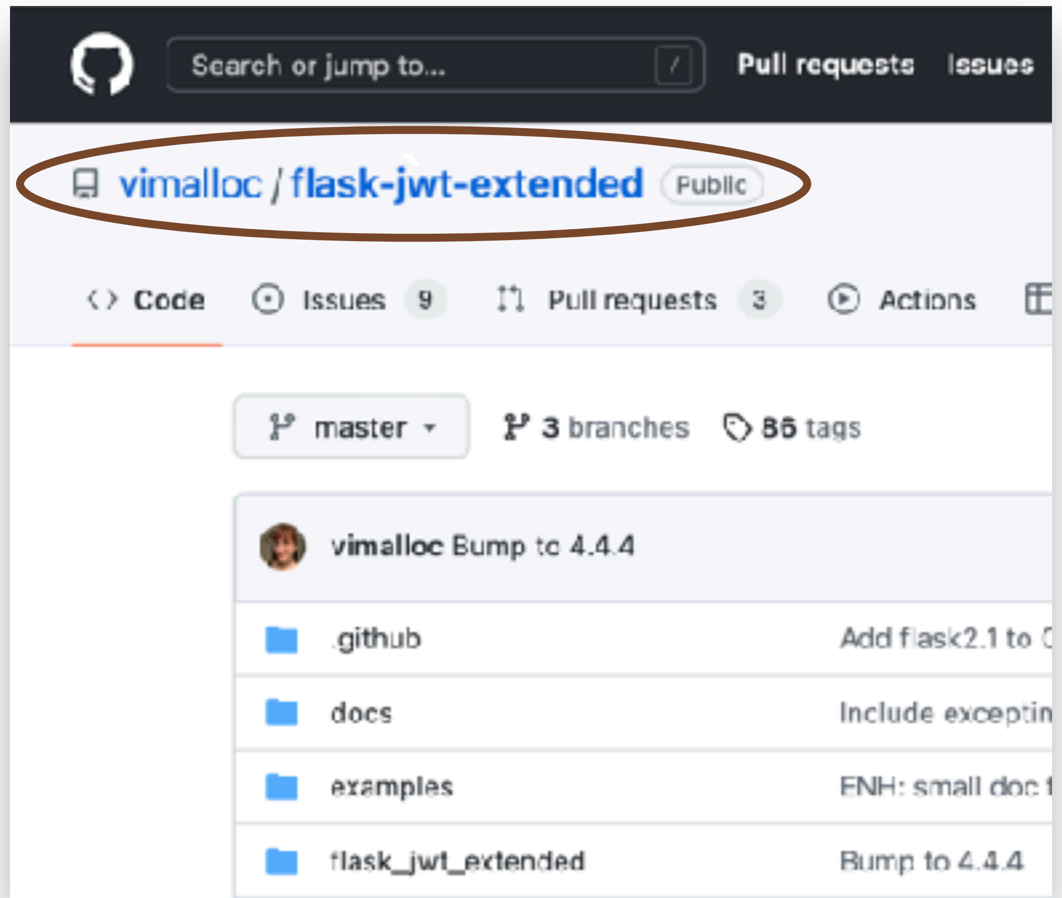


mattupstate / flask-jwt Public

Matt Wright Bump version number to 0.3.2

docs	Bump version nu
example	Use PyJWT
flask_jwt	Bump version nu
scripts	Add release scri

fork

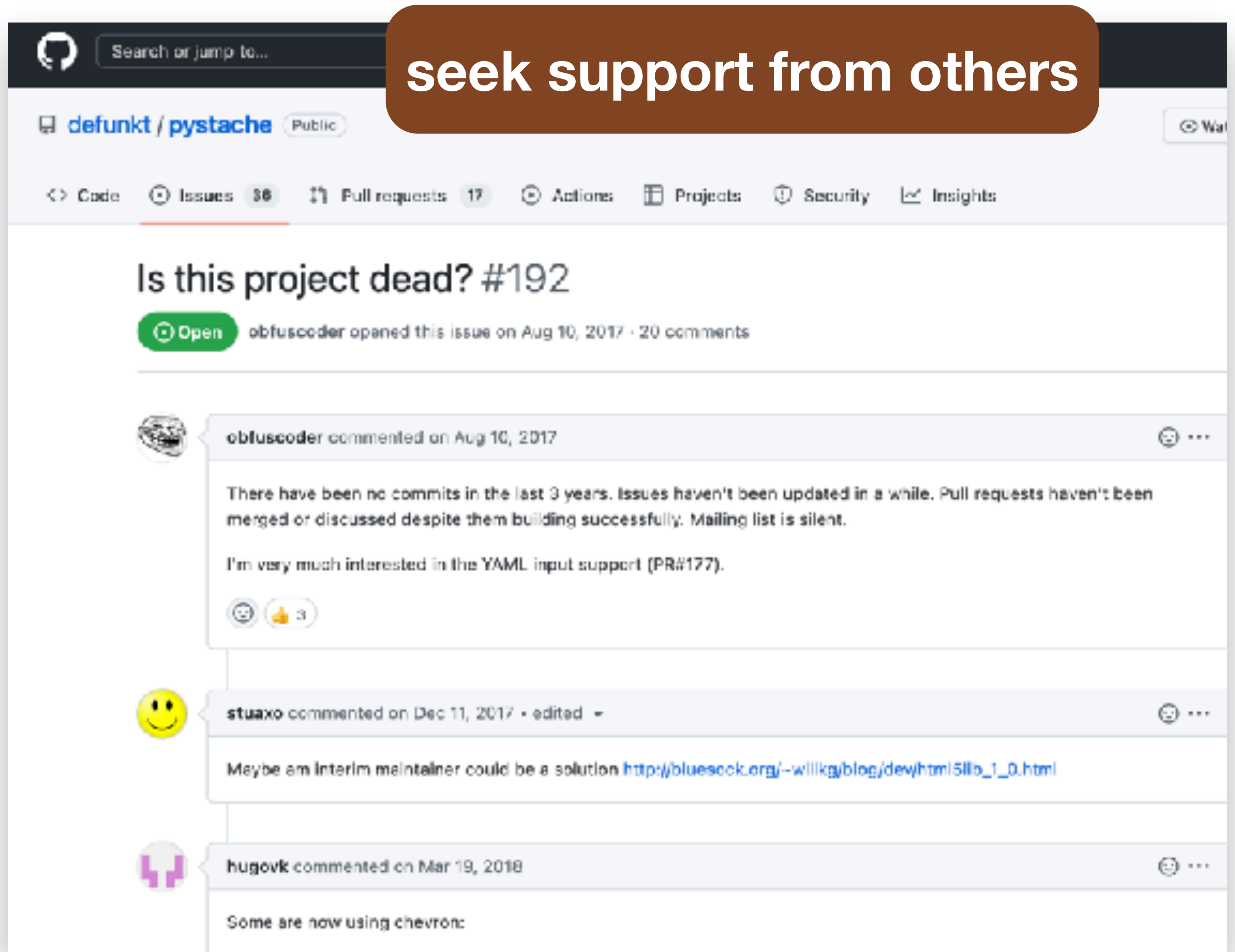


vimalloc / flask-jwt-extended Public

vimalloc Bump to 4.4.4

github	Add flask2.1 to C
docs	Include exceptin
examples	ENH: small doc f
flask_jwt_extended	Bump to 4.4.4

seek support from others



defunkt / pystache Public

Is this project dead? #192

obfuscoder opened this issue on Aug 10, 2017 · 20 comments

obfuscoder commented on Aug 10, 2017

There have been no commits in the last 3 years. Issues haven't been updated in a while. Pull requests haven't been merged or discussed despite them building successfully. Mailing list is silent.

I'm very much interested in the YAML input support (PR#177).

👍 3

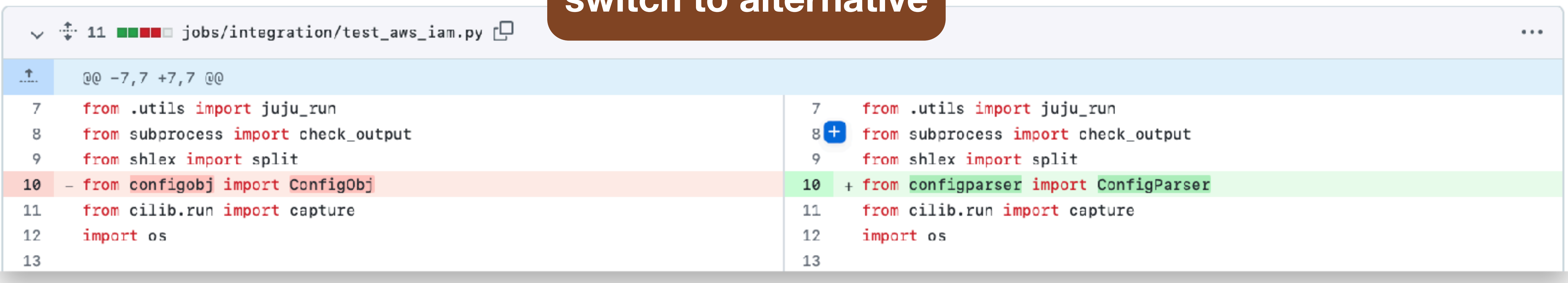
stuaxo commented on Dec 11, 2017 • edited

Maybe am interim maintainer could be a solution http://bluesock.org/~willkg/blog/dev/html5lib_1_0.html

hugovk commented on Mar 19, 2018

Some are now using chevron:

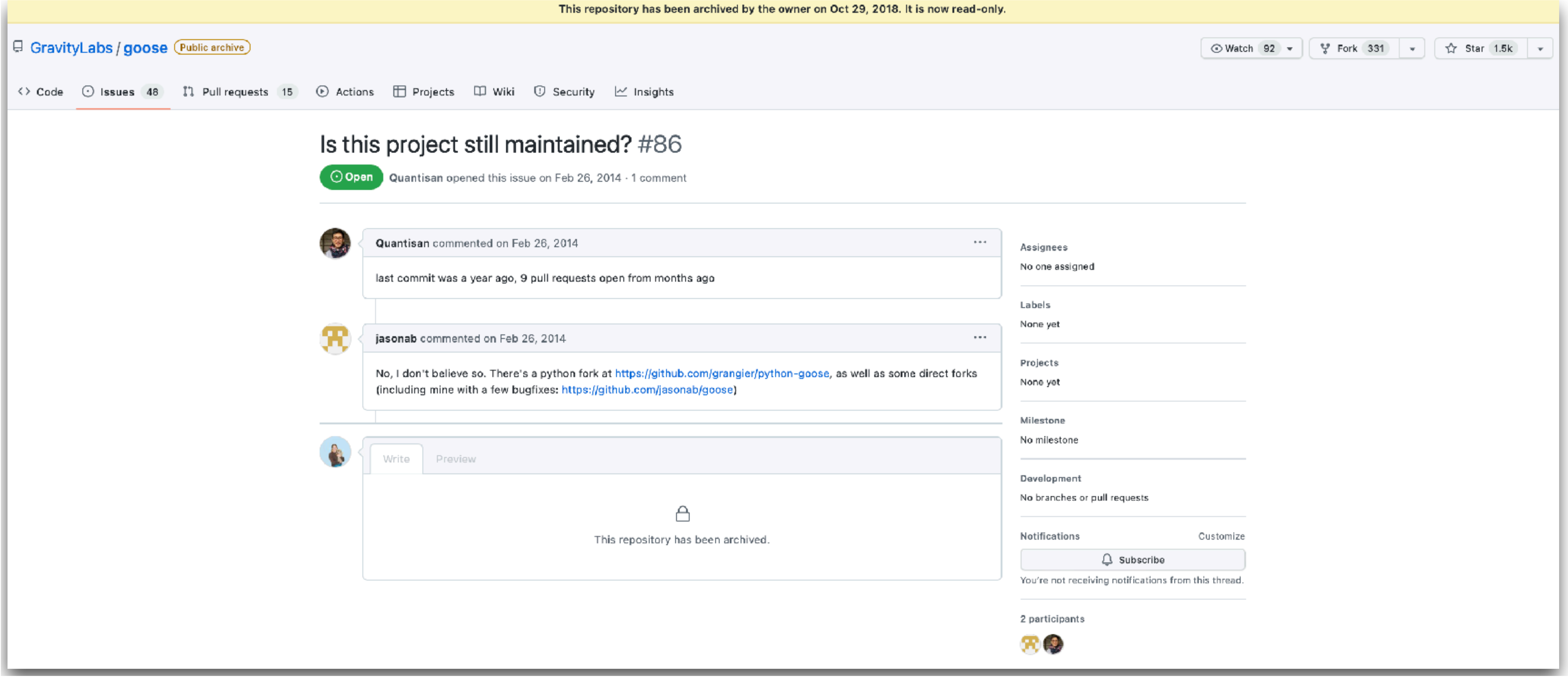
switch to alternative



jobs/integration/test_aws_iam.py

```
@@ -7,7 +7,7 @@
  from .utils import juju_run                      from .utils import juju_run
  from subprocess import check_output              from subprocess import check_output
  from shlex import split                          from shlex import split
- from configobj import ConfigObj                + from configparser import ConfigParser
  from cilib.run import capture                    from cilib.run import capture
  import os                                        import os
```

Common theme: Interviewees benefitted from the actions of others



This repository has been archived by the owner on Oct 29, 2018. It is now read-only.

GravityLabs / goose Public archive

Is this project still maintained? #86

Open Quantisan opened this issue on Feb 26, 2014 · 1 comment

Quantisan commented on Feb 26, 2014

last commit was a year ago, 9 pull requests open from months ago

jasonab commented on Feb 26, 2014

No, I don't believe so. There's a python fork at https://github.com/grangier/python-goose, as well as some direct forks (including mine with a few bugfixes: https://github.com/jasonab/goose)

This repository has been archived.

Assignees
No one assigned

Labels
None yet

Projects
None yet

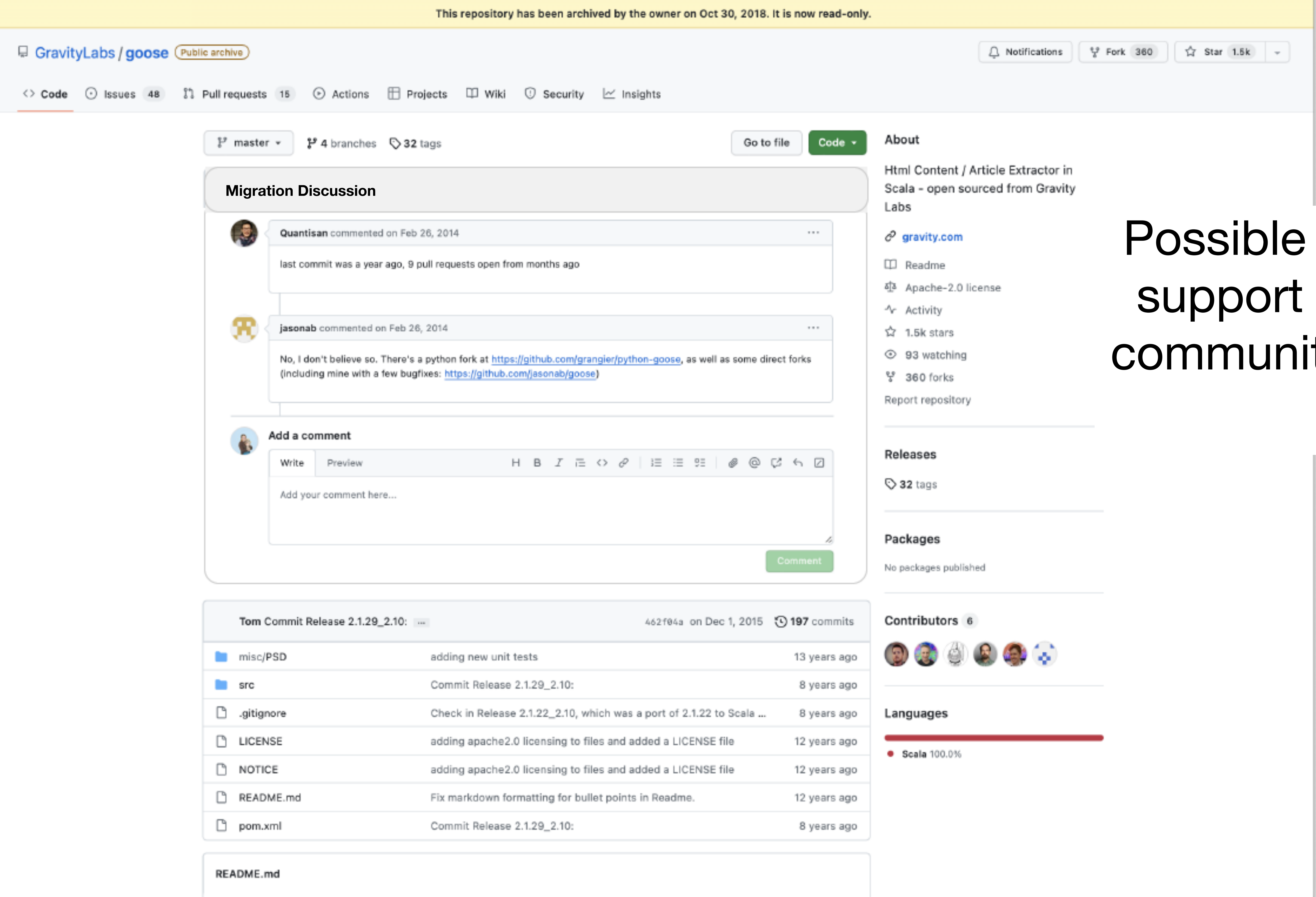
Milestone
No milestone

Development
No branches or pull requests

Notifications Customize

You're not receiving notifications from this thread.

2 participants



This repository has been archived by the owner on Oct 30, 2018. It is now read-only.

GravityLabs / goose Public archive

Migration Discussion

Quantisan commented on Feb 26, 2014

last commit was a year ago, 9 pull requests open from months ago

jasonab commented on Feb 26, 2014

No, I don't believe so. There's a python fork at https://github.com/grangier/python-goose, as well as some direct forks (including mine with a few bugfixes: https://github.com/jasonab/goose)

Add a comment

Tom Commit Release 2.1.29_2.10: ... 462f04a on Dec 1, 2015 197 commits

misc/PSD	adding new unit tests	13 years ago
src	Commit Release 2.1.29_2.10:	8 years ago
.gitignore	Check in Release 2.1.22_2.10, which was a port of 2.1.22 to Scala ...	8 years ago
LICENSE	adding apache2.0 licensing to files and added a LICENSE file	12 years ago
NOTICE	adding apache2.0 licensing to files and added a LICENSE file	12 years ago
README.md	Fix markdown formatting for bullet points in Readme.	12 years ago
pom.xml	Commit Release 2.1.29_2.10:	8 years ago

README.md

About

Html Content / Article Extractor in Scala - open sourced from Gravity Labs

gravity.com

Readme

Apache-2.0 license

Activity

1.5k stars

93 watching

360 forks

Report repository

Releases

32 tags

Packages

No packages published

Contributors 6

Languages

Scala 100.0%

Possible solution to support creation of community-oriented solutions?

Part 1 Summary:

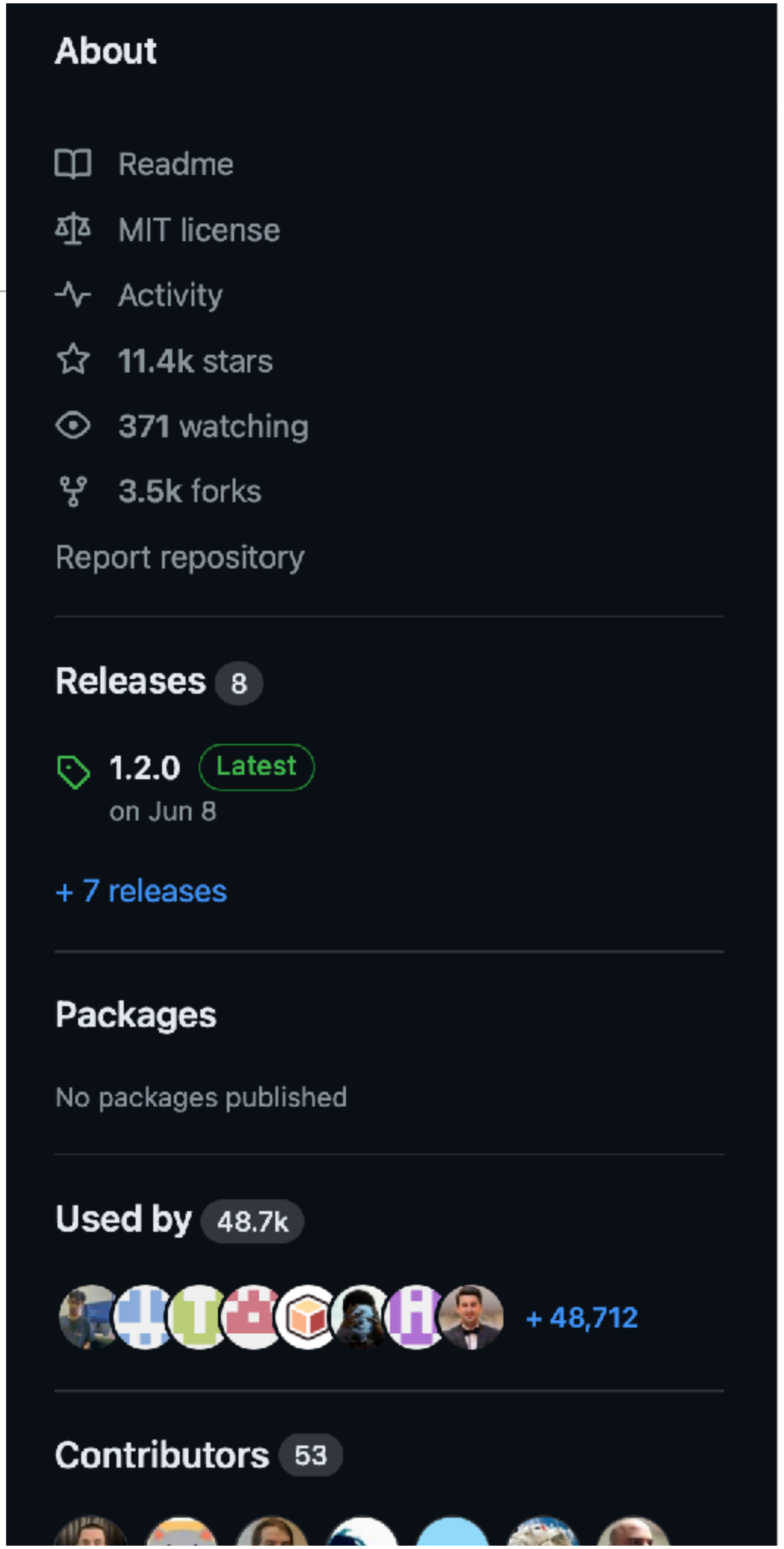
Every time a project becomes abandoned, or we think it might be abandoned, we feel like we're winging it.

We feel like we're dealing with it for the first time

- PID4

Part 2: Repository Mining

28,100 npm packages out of 1M+ in 2020 had at least one month with 10,000+ downloads



About

Readme

MIT license

Activity

11.4k stars

371 watching

3.5k forks

Report repository

Releases 8

1.2.0 Latest
on Jun 8

+ 7 releases

Packages

No packages published

Used by 48.7k

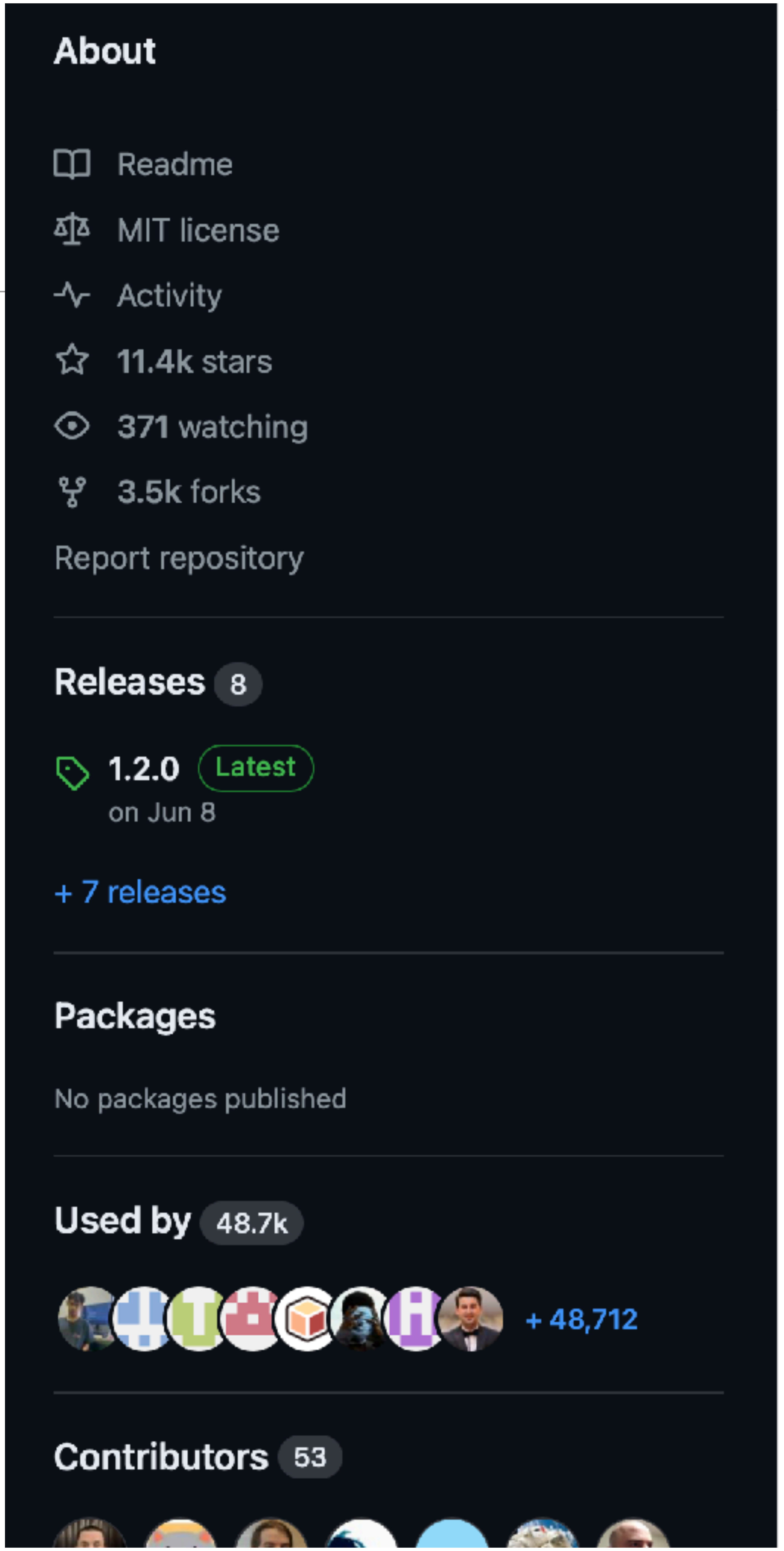
+ 48,712

Contributors 53

28,100 npm packages out of 1M+ in 2020 had at least one month with 10,000+ downloads

15% (4,108) became abandoned

Observation window: Jan 2015 to Dec 2020



About

Readme

MIT license

Activity

11.4k stars

371 watching

3.5k forks

Report repository

Releases 8

1.2.0 Latest

on Jun 8

+ 7 releases

Packages

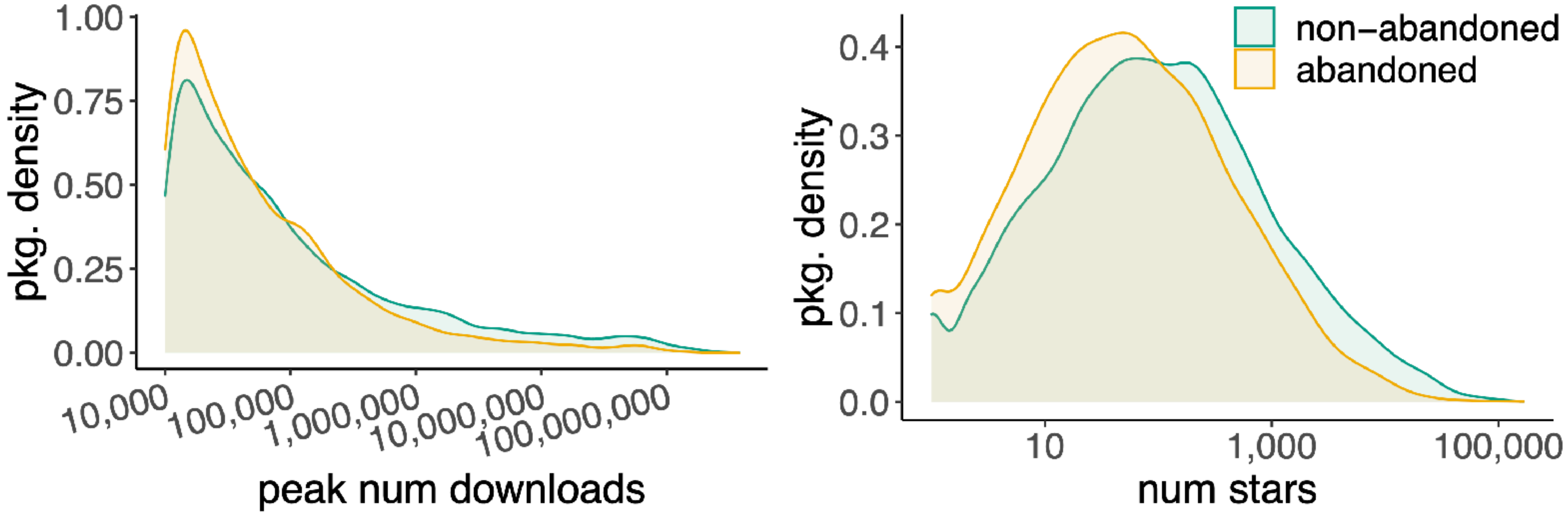
No packages published

Used by 48.7k

+ 48,712

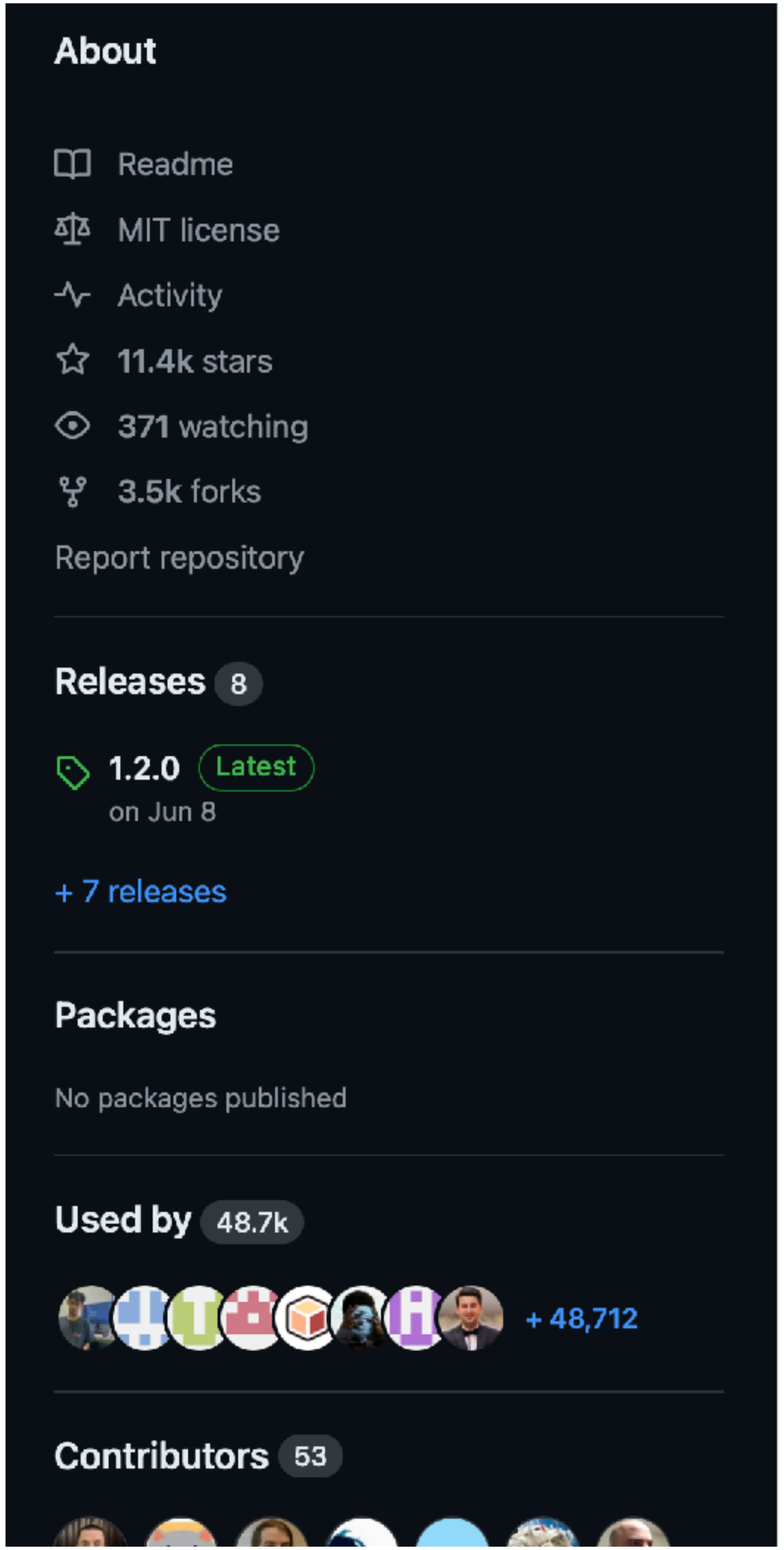
Contributors 53

The distributions of peak download and current star counts for both abandoned and non-abandoned packages are similar.



The abandoned projects impacted

~280k+ downstreams on GitHub



About

Readme

MIT license

Activity

11.4k stars

371 watching

3.5k forks

Report repository

Releases 8

1.2.0 Latest
on Jun 8

+ 7 releases

Packages

No packages published

Used by 48.7k

+ 48,712

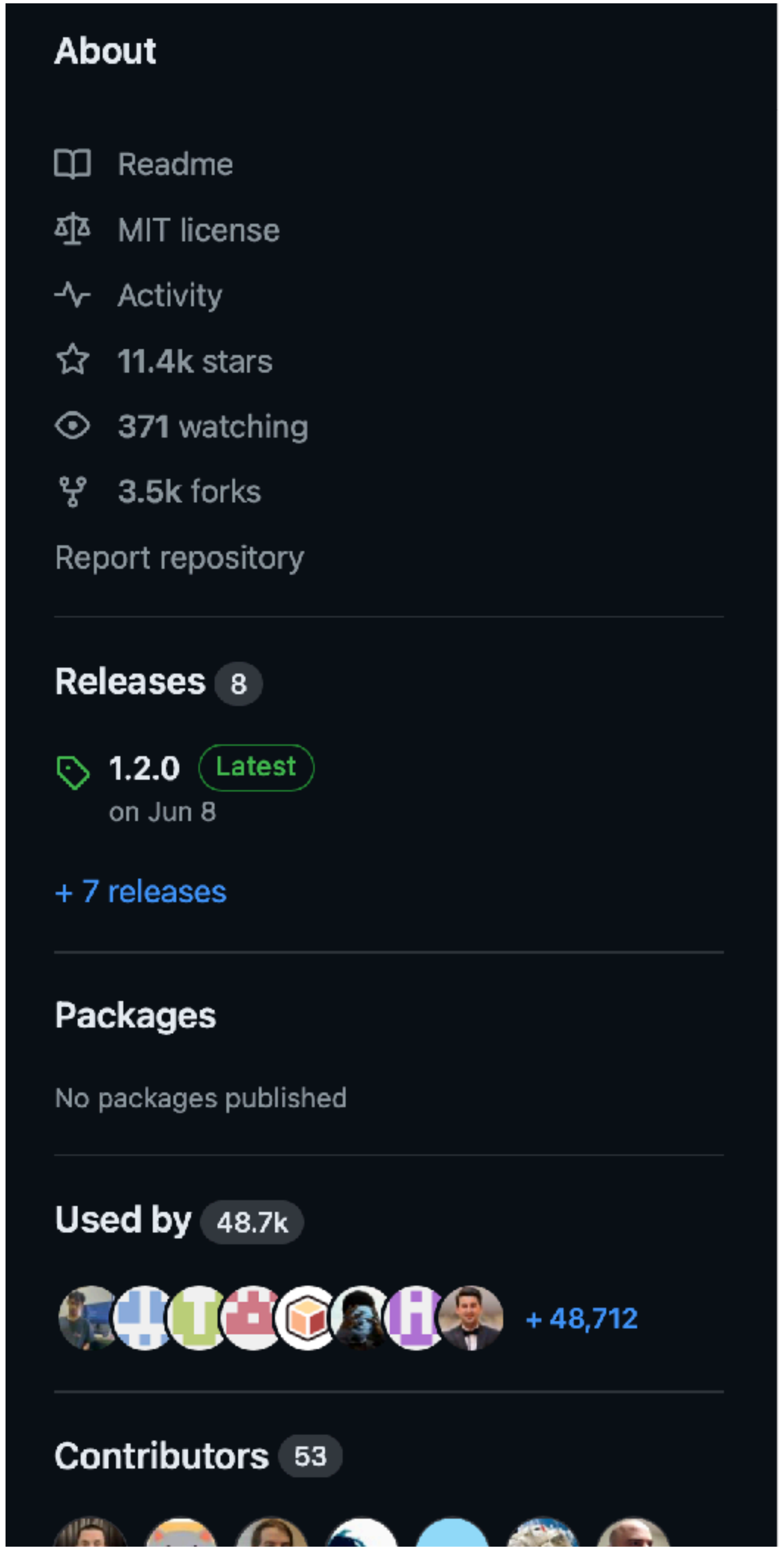
Contributors 53

The abandoned projects impacted

~280k+ downstreams on GitHub

of which

~78k+ were still active at the time



About

Readme

MIT license

Activity

11.4k stars

371 watching

3.5k forks

Report repository

Releases 8

1.2.0 Latest
on Jun 8

+ 7 releases

Packages

No packages published

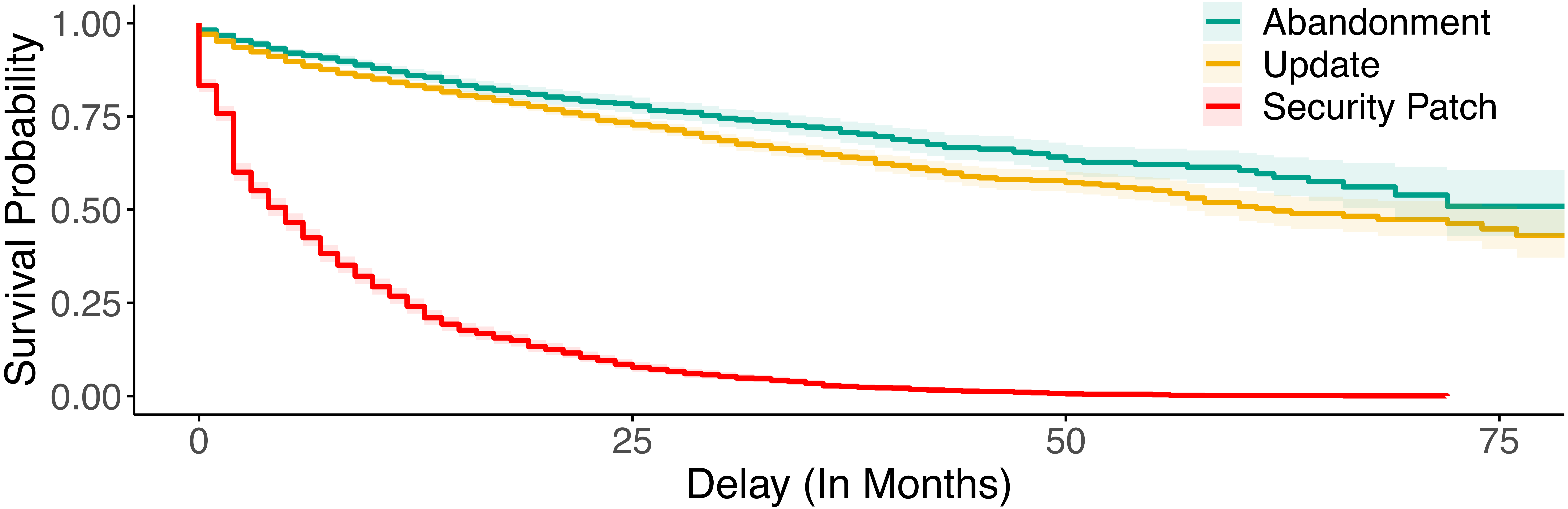
Used by 48.7k

+ 48,712

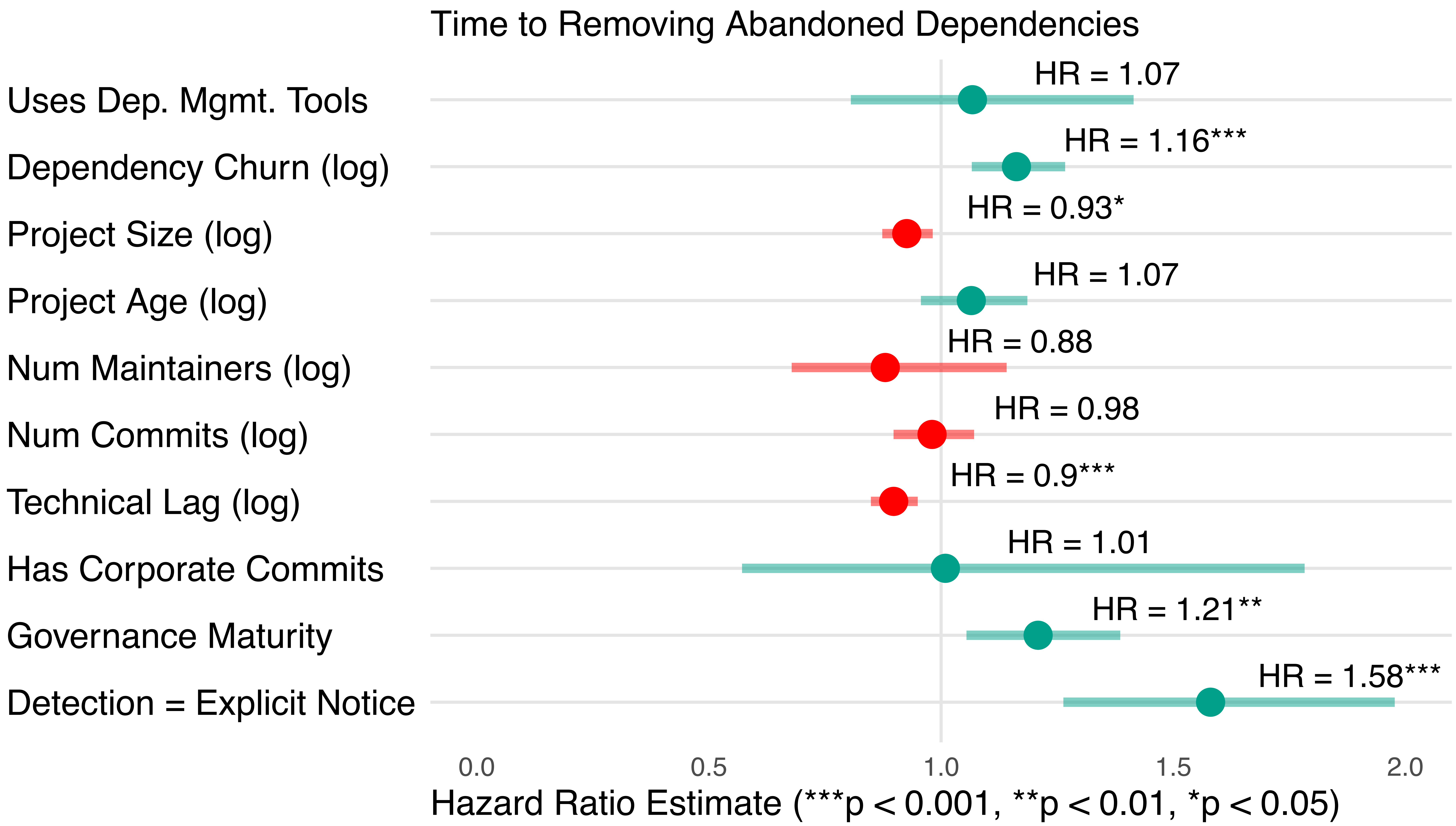
Contributors 53

How much do people downstream react?

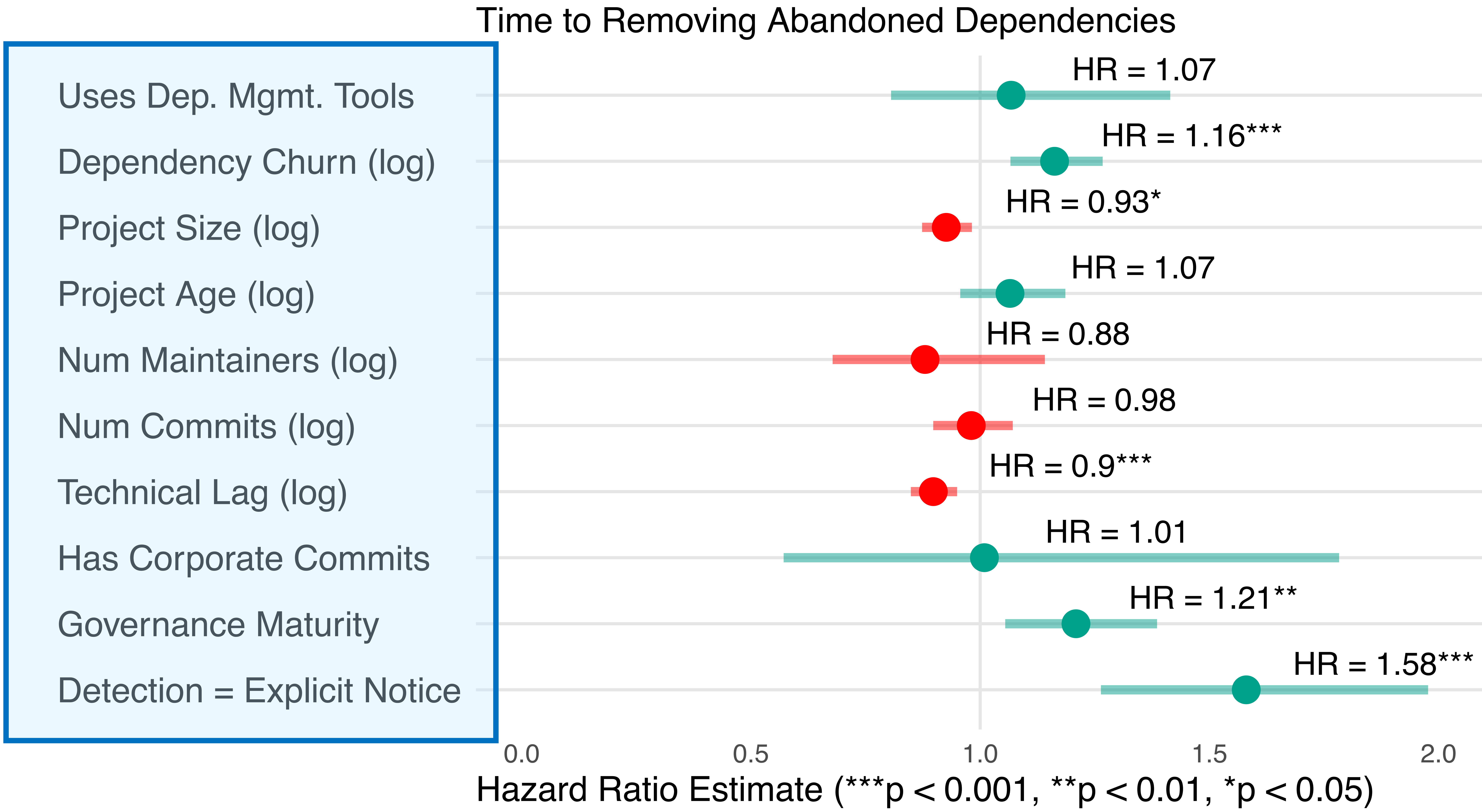
The rate of removing abandoned dependencies is similar to random dependency updates, and slower than security patch updates.



Which factors correlate with downstream projects reacting faster?

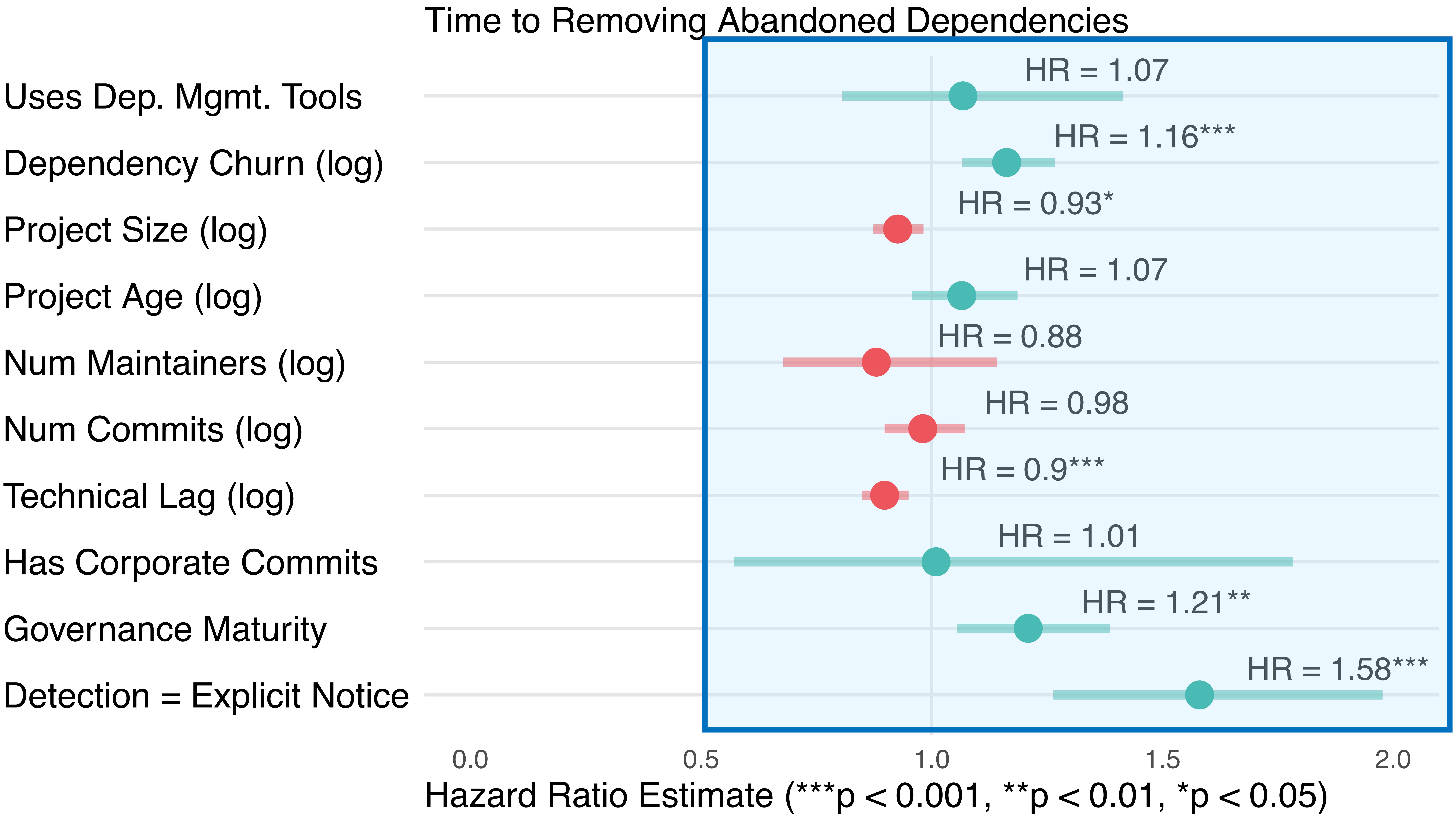

Time to Removing Abandoned Dependencies
Uses Dep. Mgmt. Tools
HR = 1.07
Dependency Churn (log)
HR = 1.16***
Project Size (log)
HR = 0.93*
Project Age (log)
HR = 1.07
Num Maintainers (log)
HR = 0.88
Num Commits (log)
HR = 0.98
Technical Lag (log)
HR = 0.9***
Has Corporate Commits
HR = 1.01
Governance Maturity
HR = 1.21**
Detection = Explicit Notice
HR = 1.58***
0.0
0.5
1.0
1.5
2.0
Hazard Ratio Estimate (***p < 0.001, **p < 0.01, *p < 0.05)

Factors

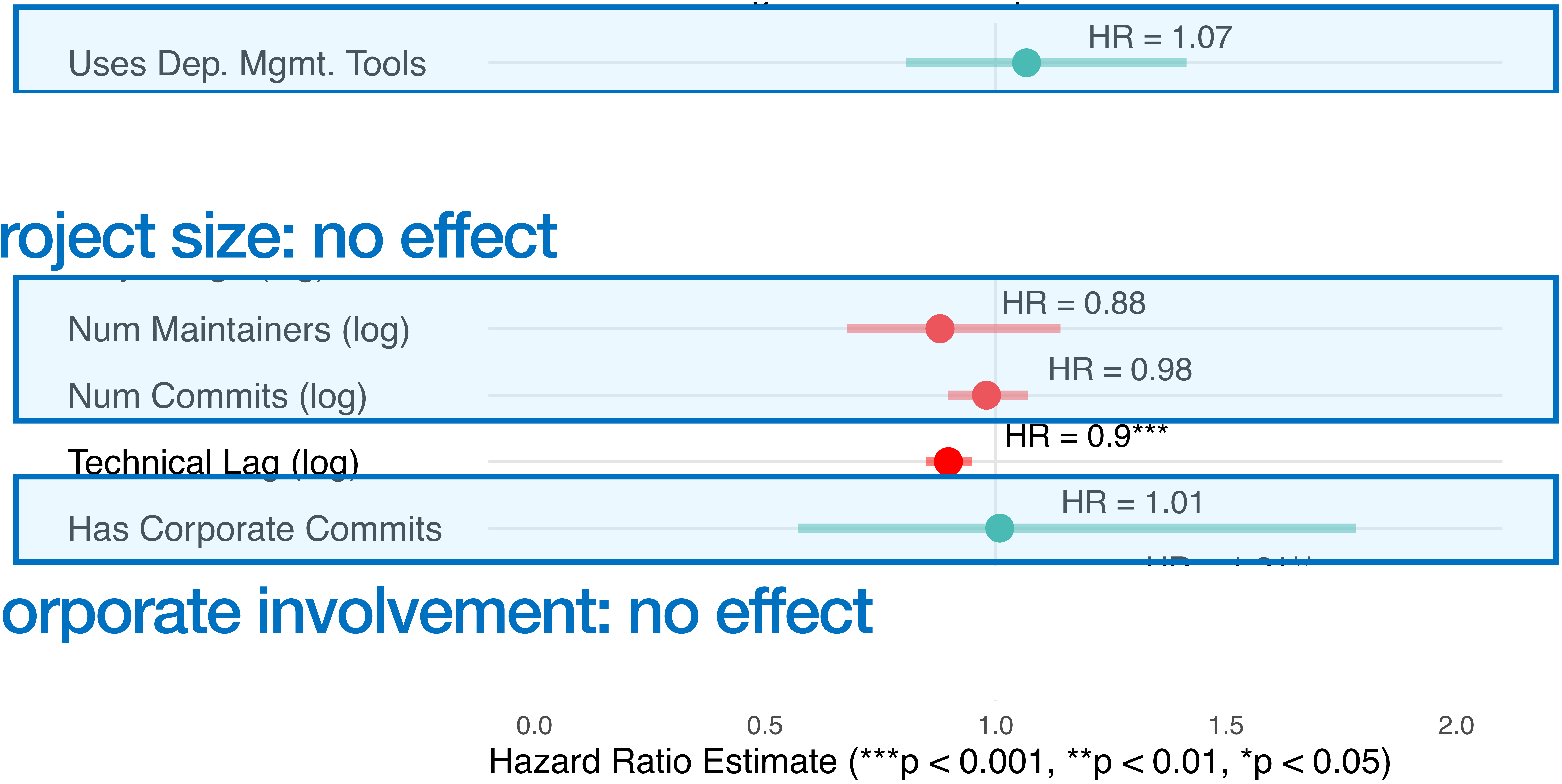


Time to Removing Abandoned Dependencies
Uses Dep. Mgmt. Tools
HR = 1.07
Dependency Churn (log)
HR = 1.16***
Project Size (log)
HR = 0.93*
Project Age (log)
HR = 1.07
Num Maintainers (log)
HR = 0.88
Num Commits (log)
HR = 0.98
Technical Lag (log)
HR = 0.9***
Has Corporate Commits
HR = 1.01
Governance Maturity
HR = 1.21**
Detection = Explicit Notice
HR = 1.58***
0.0
0.5
1.0
1.5
2.0
Hazard Ratio Estimate (***p < 0.001, **p < 0.01, *p < 0.05)

Magnitude of correlation

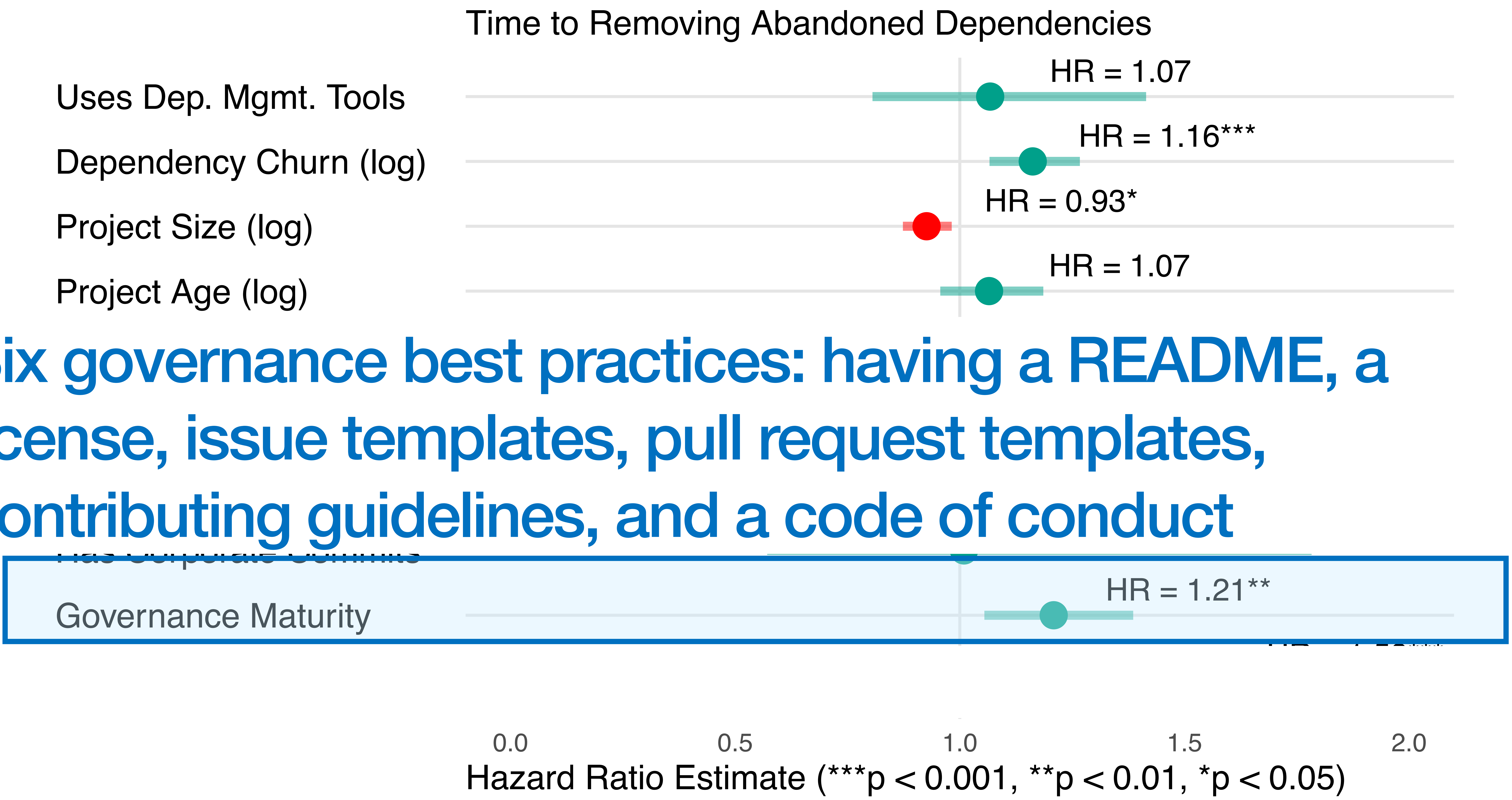


Automation: no effect



Project size: no effect

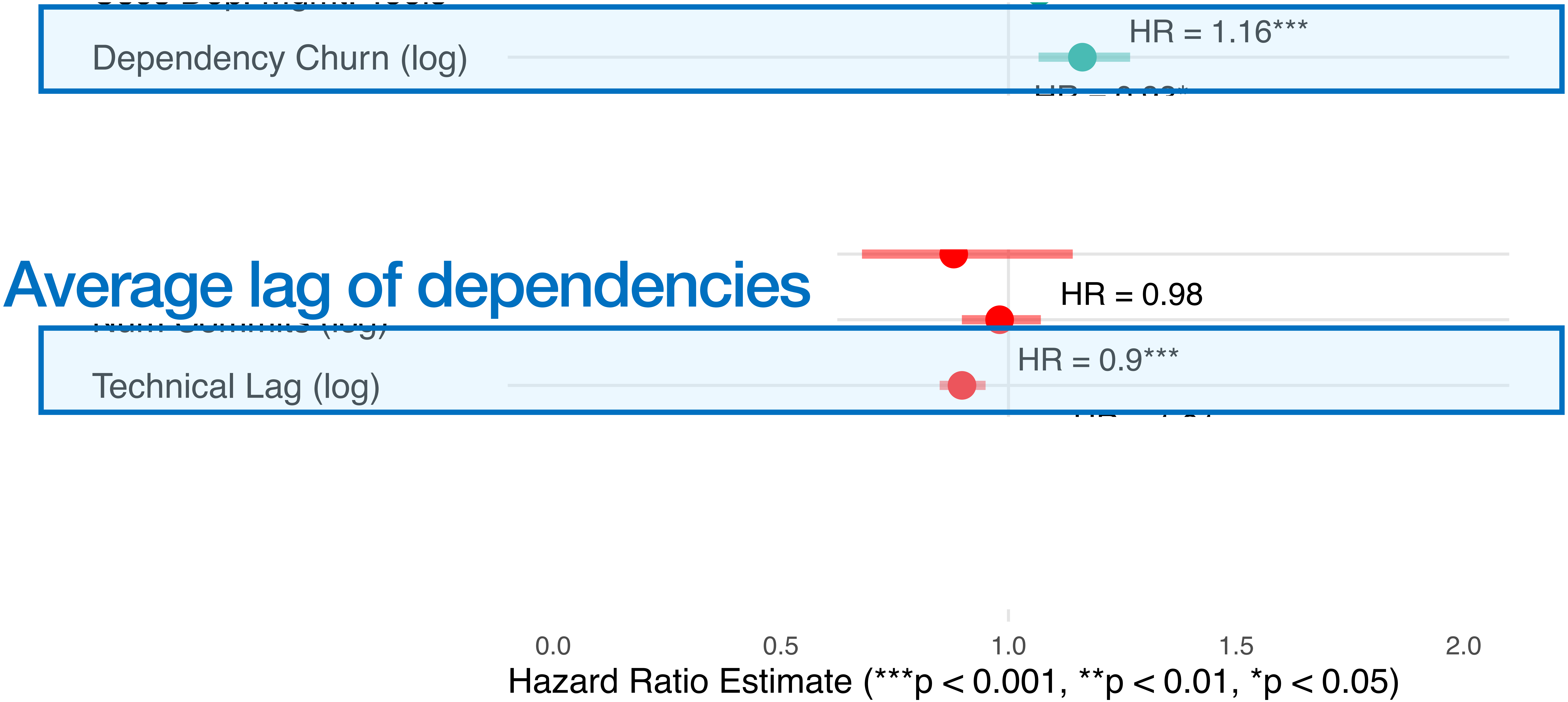
Corporate involvement: no effect

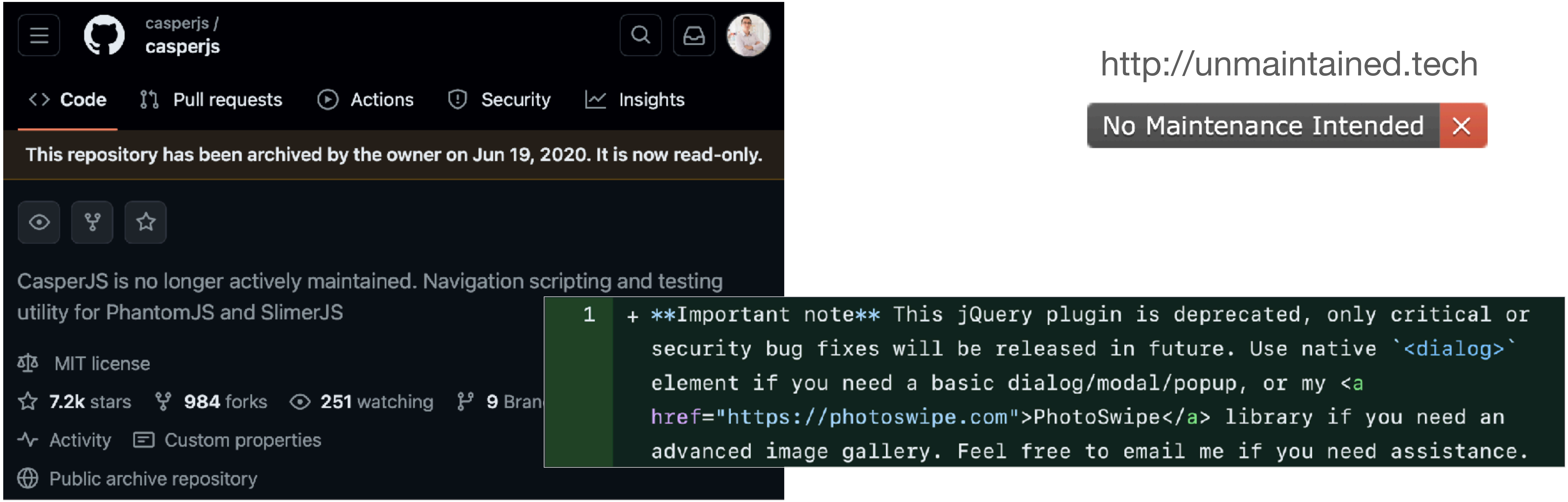


Six governance best practices: having a README, a license, issue templates, pull request templates, contributing guidelines, and a code of conduct

Updates to dependencies in the year before exposure

Average lag of dependencies





casperjs / casperjs

Code | Pull requests | Actions | Security | Insights

This repository has been archived by the owner on Jun 19, 2020. It is now read-only.

CasperJS is no longer actively maintained. Navigation scripting and testing utility for PhantomJS and SlimerJS

MIT license

7.2k stars | 984 forks | 251 watching | 9 Bran

Activity | Custom properties

Public archive repository

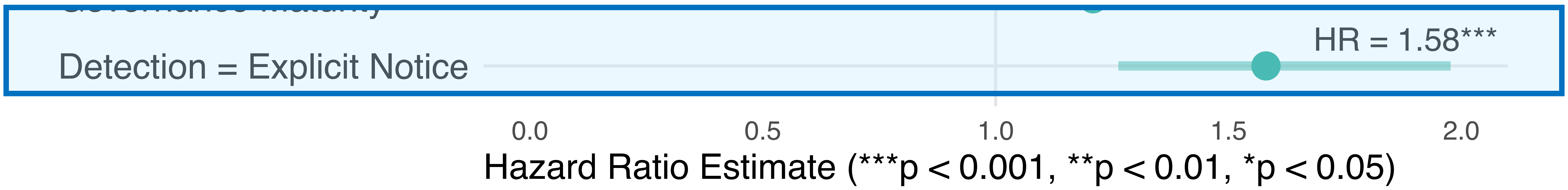
http://unmaintained.tech

No Maintenance Intended ×

```
+ **Important note** This jQuery plugin is deprecated, only critical or
  security bug fixes will be released in future. Use native `<dialog>`
  element if you need a basic dialog/modal/popup, or my <a
  href="https://photoswipe.com">PhotoSwipe</a> library if you need an
  advanced image gallery. Feel free to email me if you need assistance.
```

Strongest effect: Explicit notice of abandonment

(Github archive flag, no-maintenance-intended badge, other mention in README)



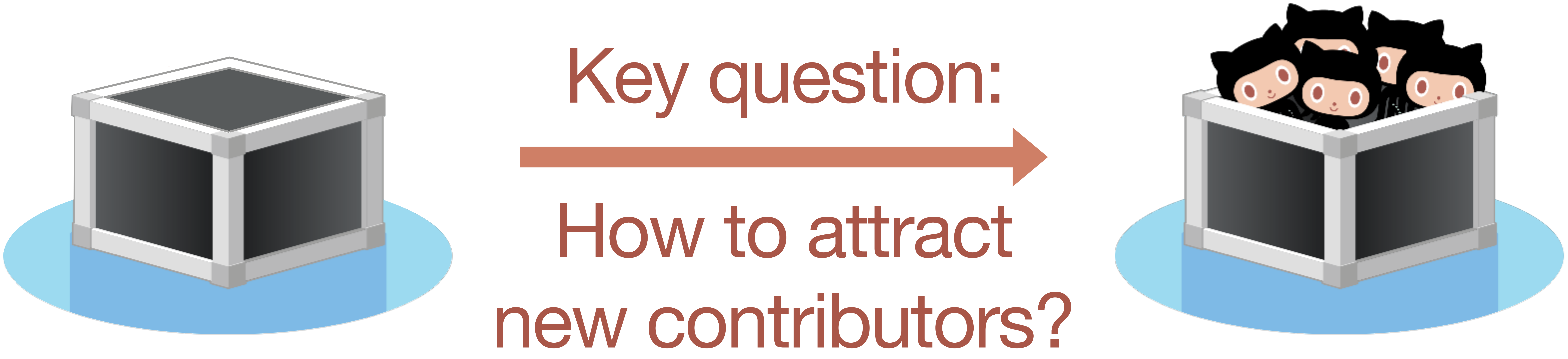
Conclusion:

- Abandonment, even among widely-used npm packages, is fairly common.
- It can have rippling effects, especially when considering transitive impact.
- People seem to care about abandoned dependencies (many remove them), but may not notice them. It's also unclear what to do after.
- At the very least, we recommend that:
  - Maintainers place an **explicit notice** of abandonment somewhere visible.
  - Platforms implement features to **help with migration**.
- It's time to establish best practices for **responsible sunsetting** of packages, rather than insisting on indefinite maintenance!

Labor Pools

Fang, Herbsleb, and Vasilescu, “Matching Skills, Past Collaboration, and Limited Competition: Modeling When Open-Source Projects Attract Contributors.” ESEC/FSE 2023

Strudels with sour cherry, apricot cheese, and poppy seed filling, Strudel House Cafe, Budapest, Hungary 2017



Key question:

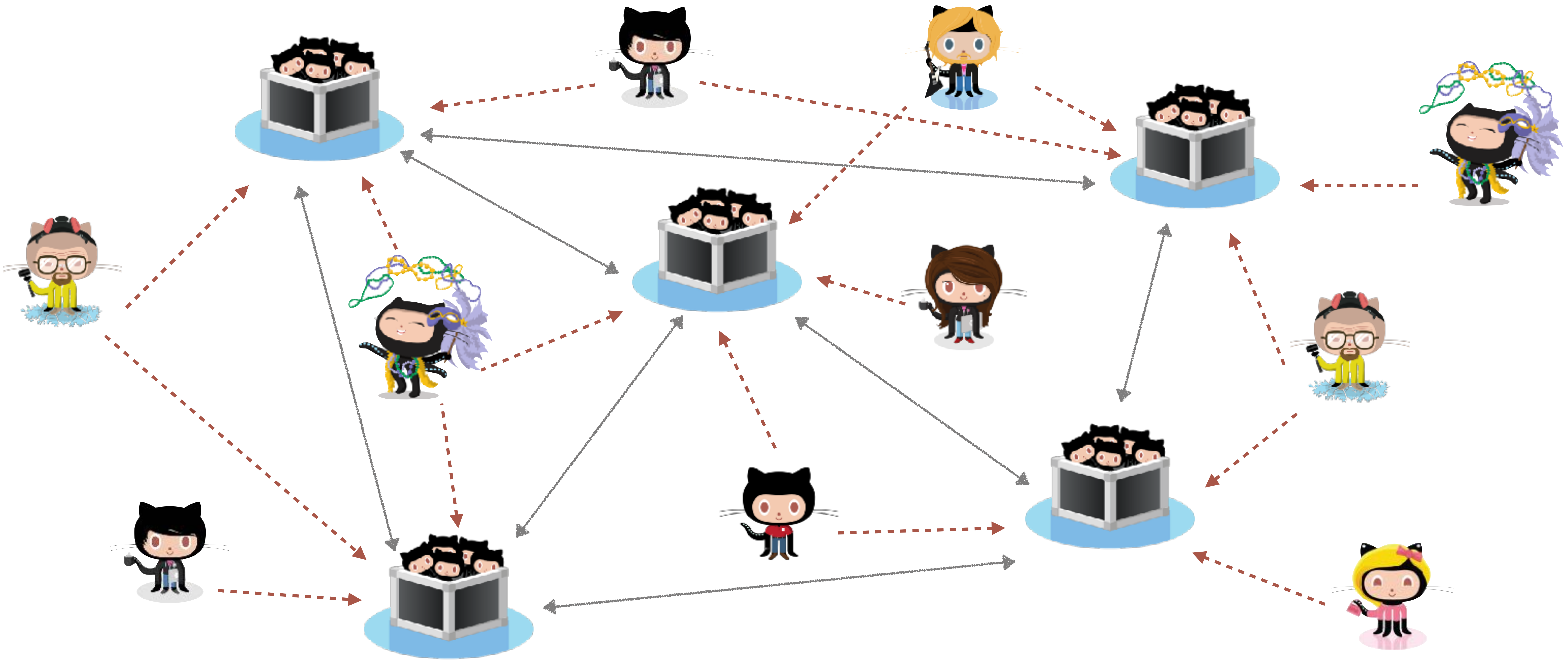
How to attract new contributors?

Many project-level factors associate with the likelihood of attracting new contributors

- Low barrier to first contribution
- Perceived welcomeness to newcomers
- Quality of README
- Current project popularity

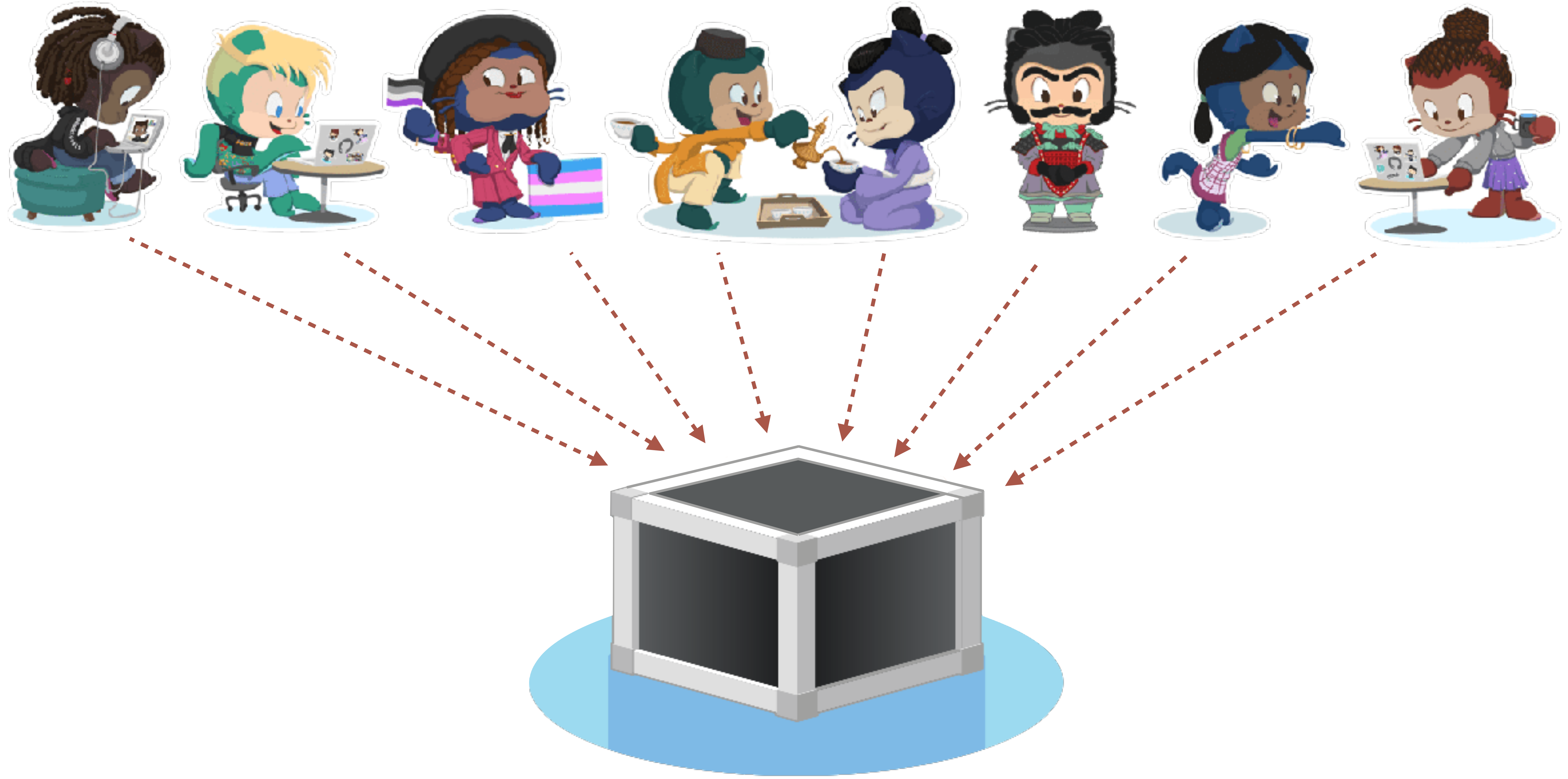
…

Open-source projects form complex networks of interdependencies!

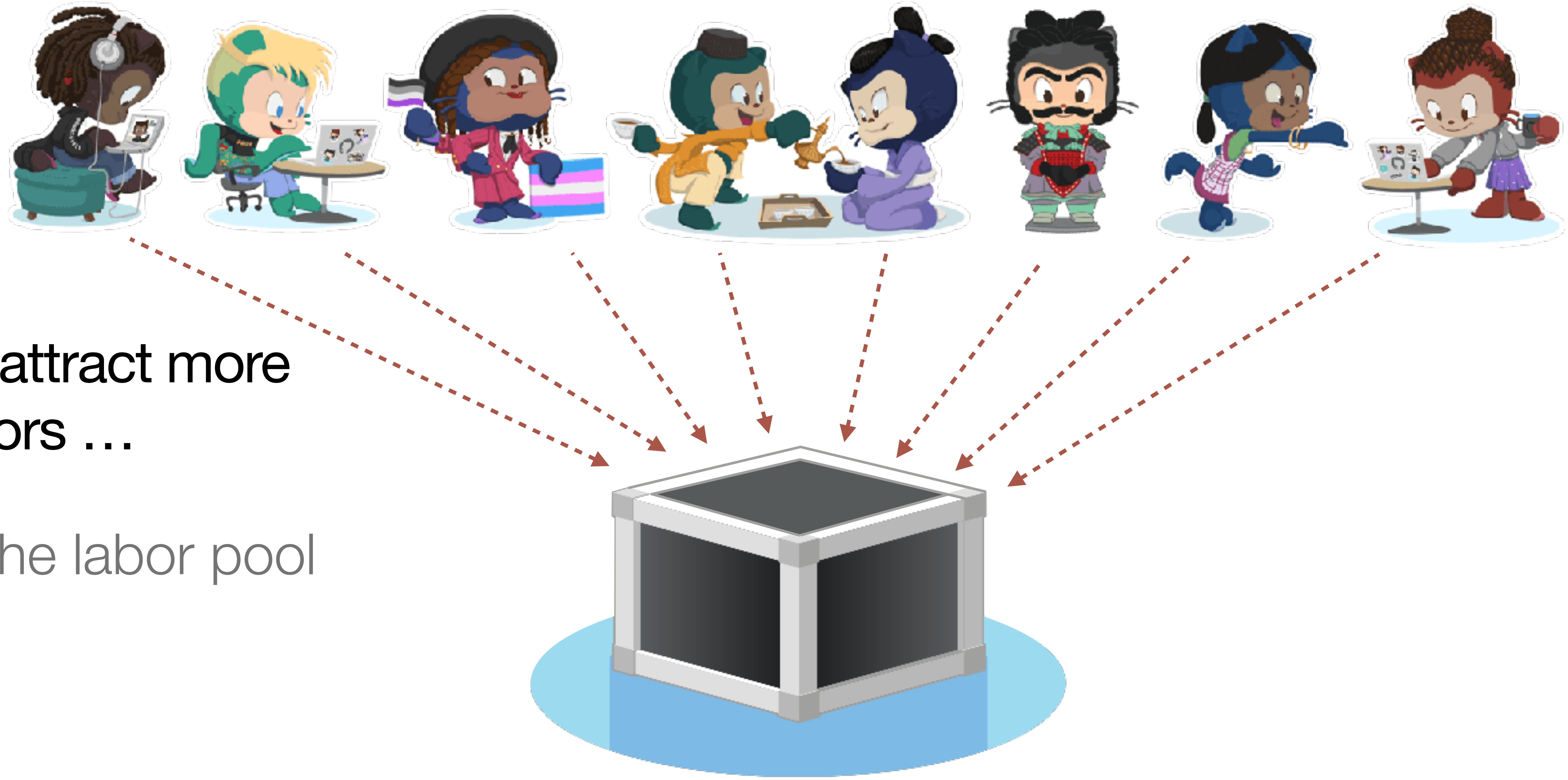


Can we measure the network effects?

New construct: a project's labor pool — the set of active participants in the overall ecosystem that the project could attempt to recruit from at a given time



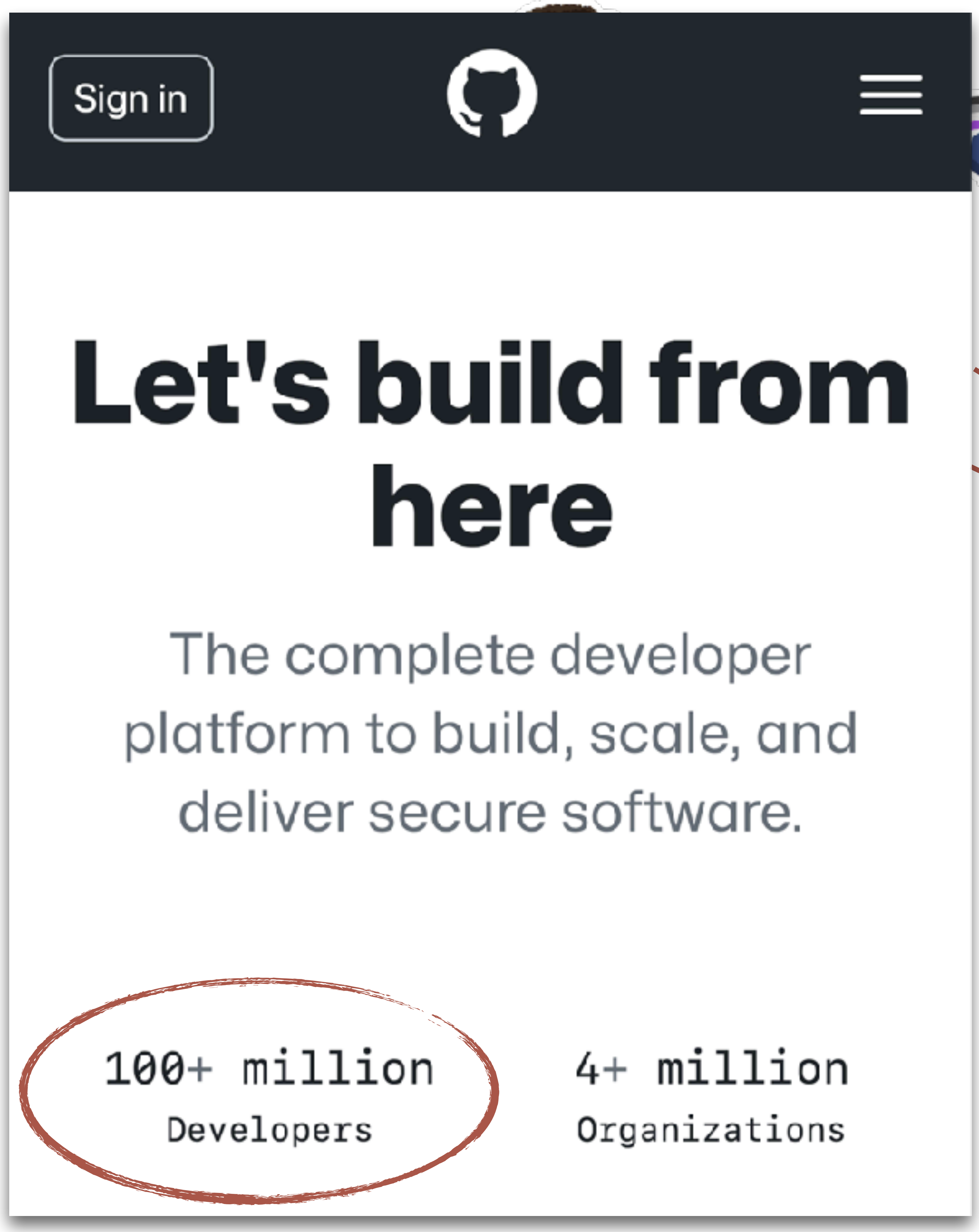
New construct: a project's labor pool — the set of active participants in the overall ecosystem that the project could attempt to recruit from at a given time



Hyp: Projects attract more new contributors …

… the larger the labor pool

New construct: a project’s labor pool — the set of active participants in the overall ecosystem that the project could attempt to recruit from at a given time



Sign in

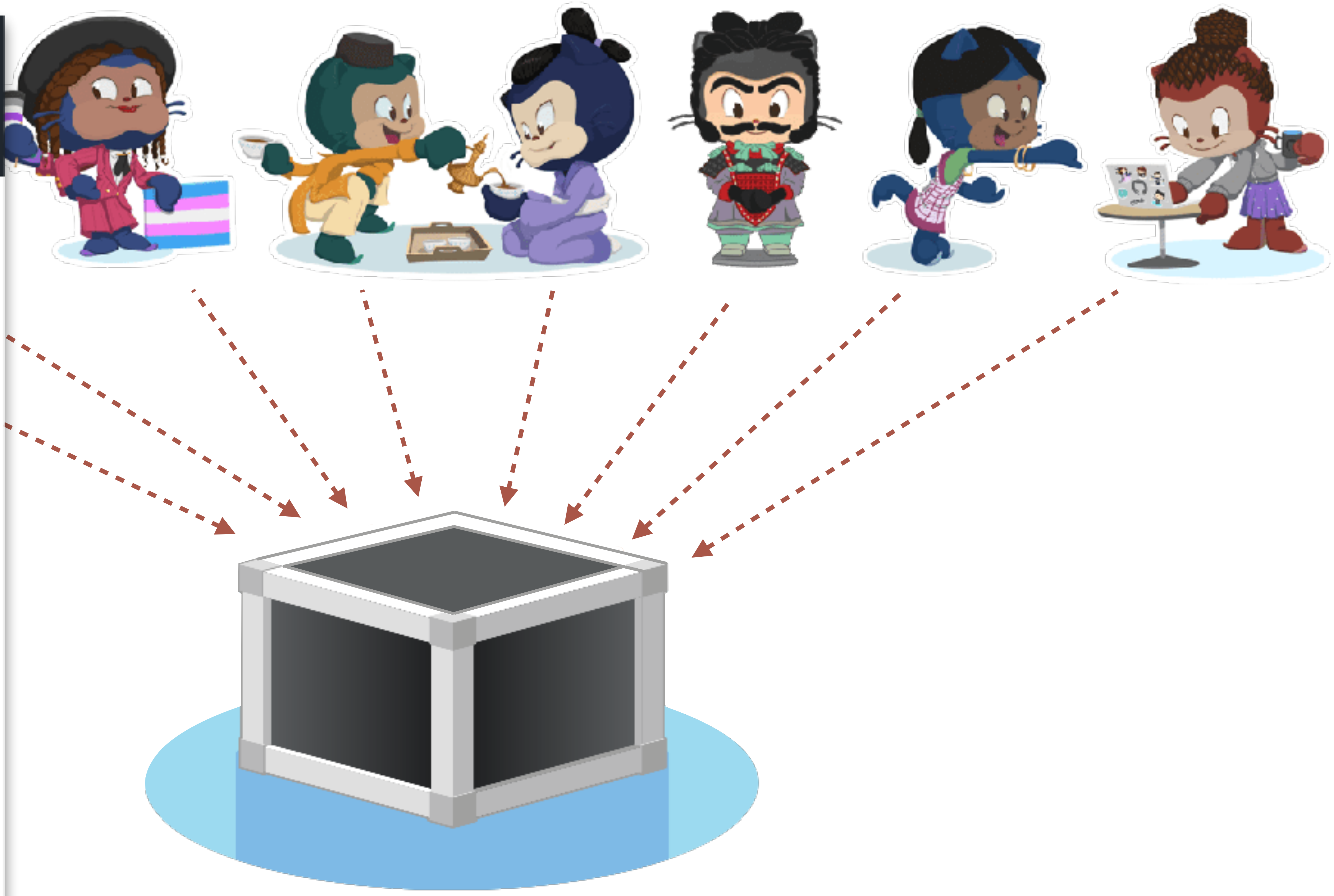
Let's build from here

The complete developer platform to build, scale, and deliver secure software.

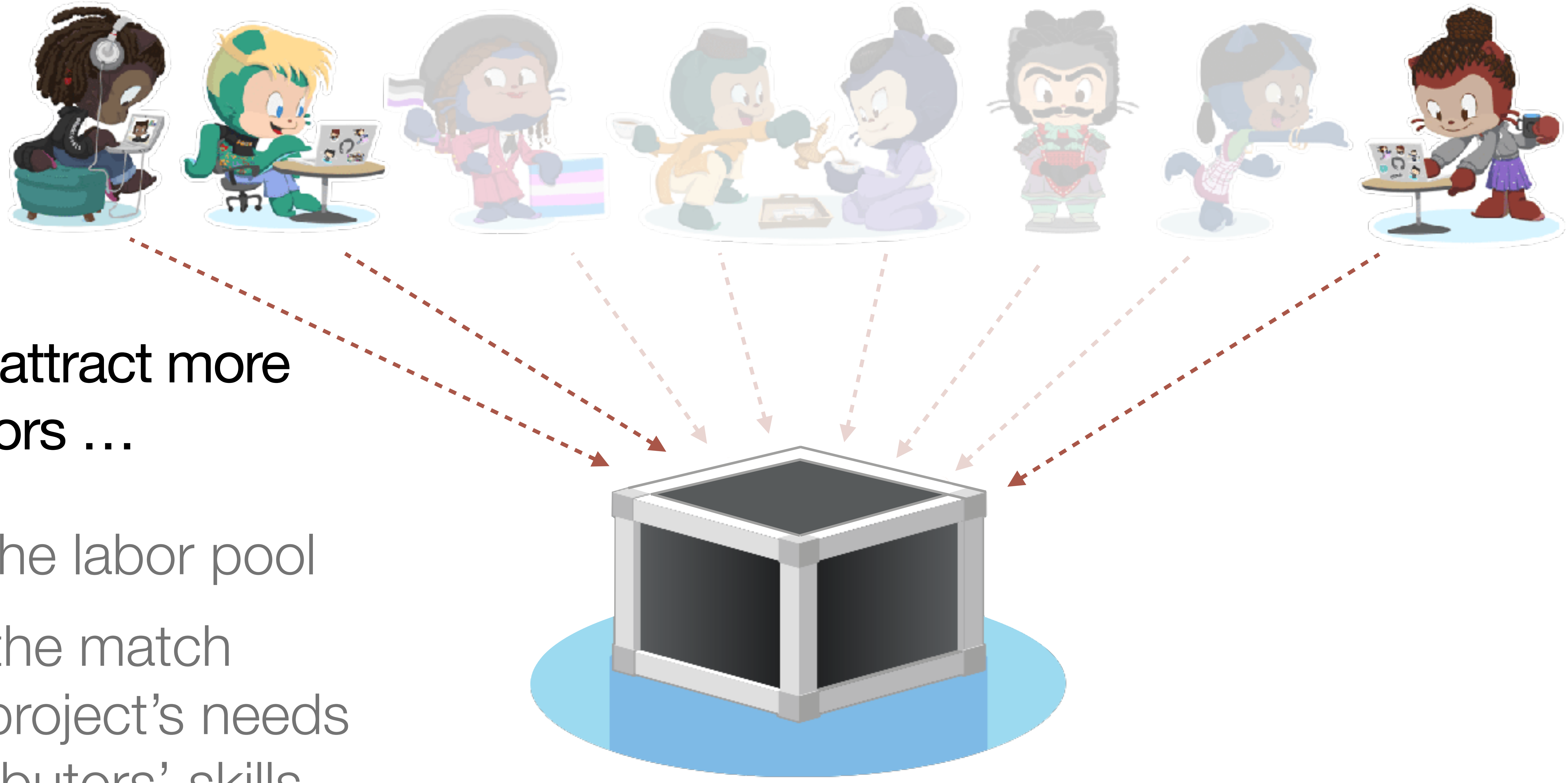
100+ million Developers

4+ million Organizations

https://github.com/about



New construct: a project’s labor pool — the set of active participants in the overall ecosystem that the project could attempt to recruit from at a given time

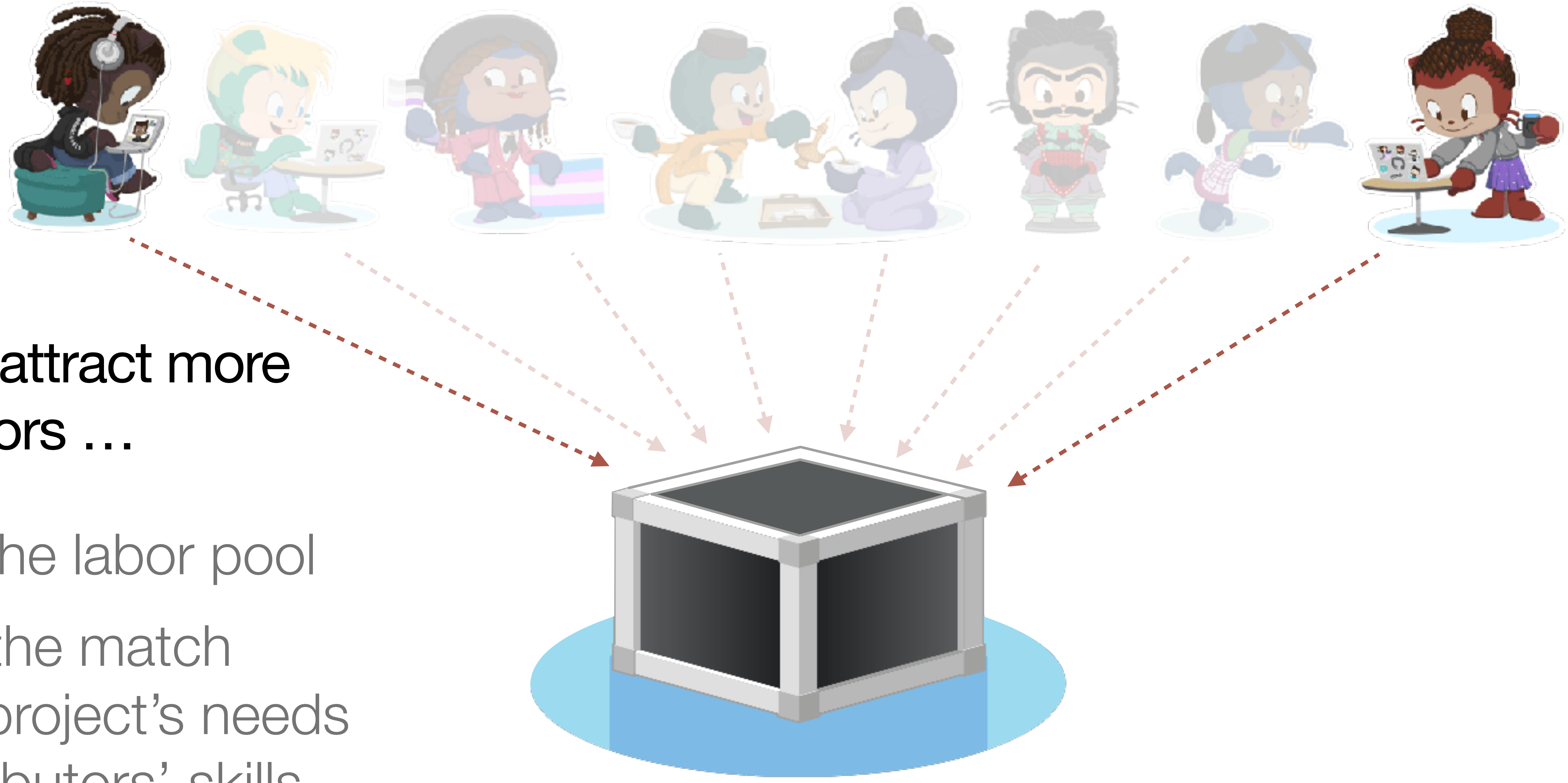


Hyp: Projects attract more new contributors …

… the larger the labor pool

… the better the match between the project’s needs and the contributors’ skills

New construct: a project's labor pool — the set of active participants in the overall ecosystem that the project could attempt to recruit from at a given time



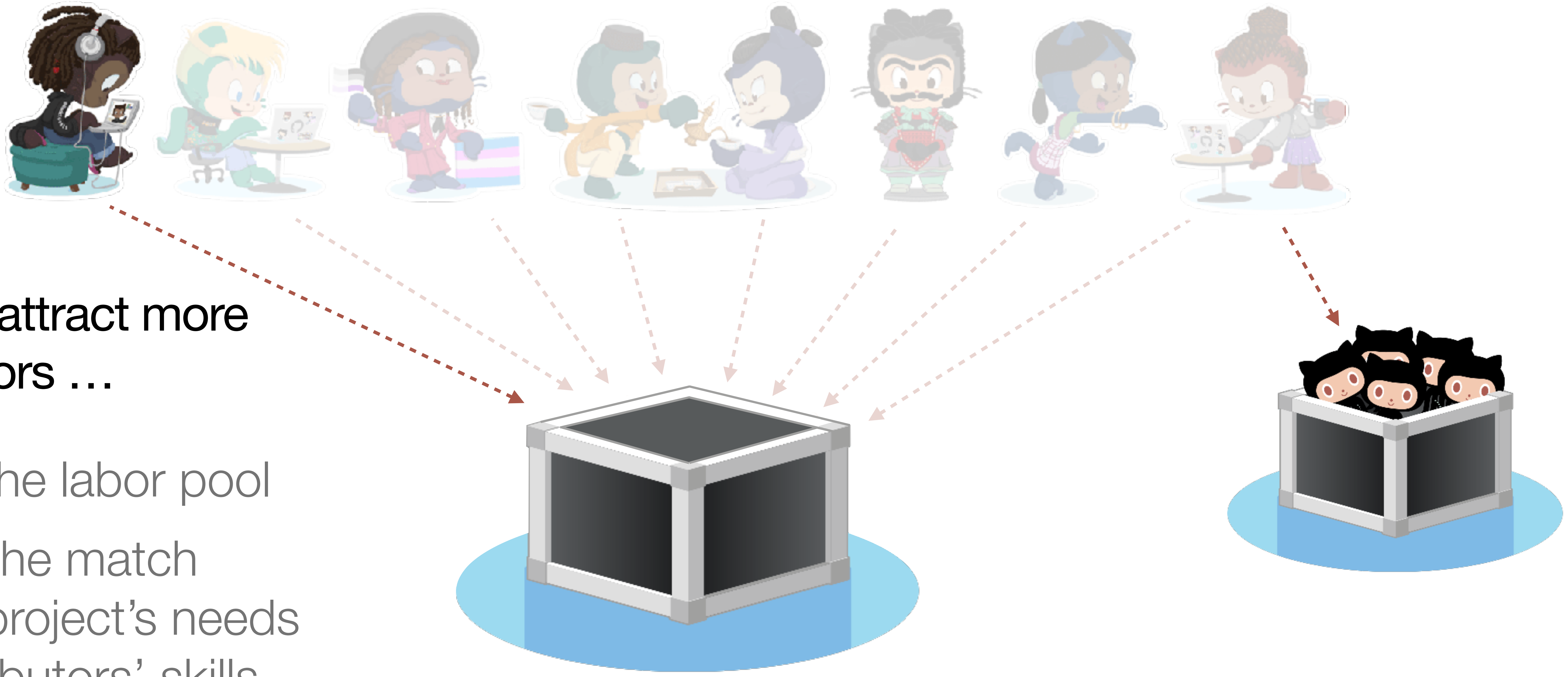
Hyp: Projects attract more new contributors …

… the larger the labor pool

… the better the match between the project's needs and the contributors' skills

… the stronger the pre-existing social connections to current project maintainers

New construct: a project’s labor pool — the set of active participants in the overall ecosystem that the project could attempt to recruit from at a given time



Hyp: Projects attract more new contributors …

… the larger the labor pool

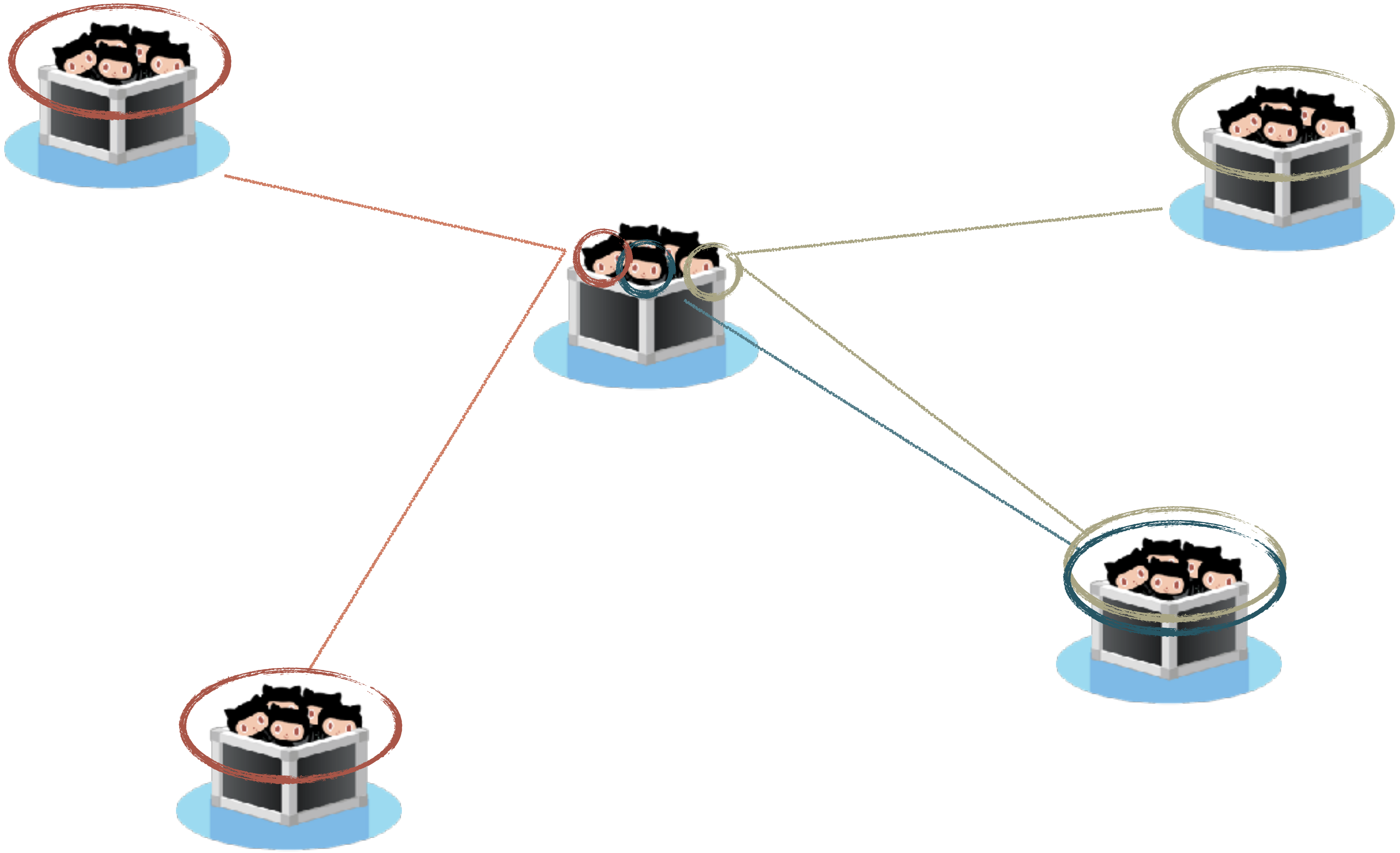
… the better the match between the project’s needs and the contributors’ skills

… the stronger the pre-existing social connections to current project maintainers

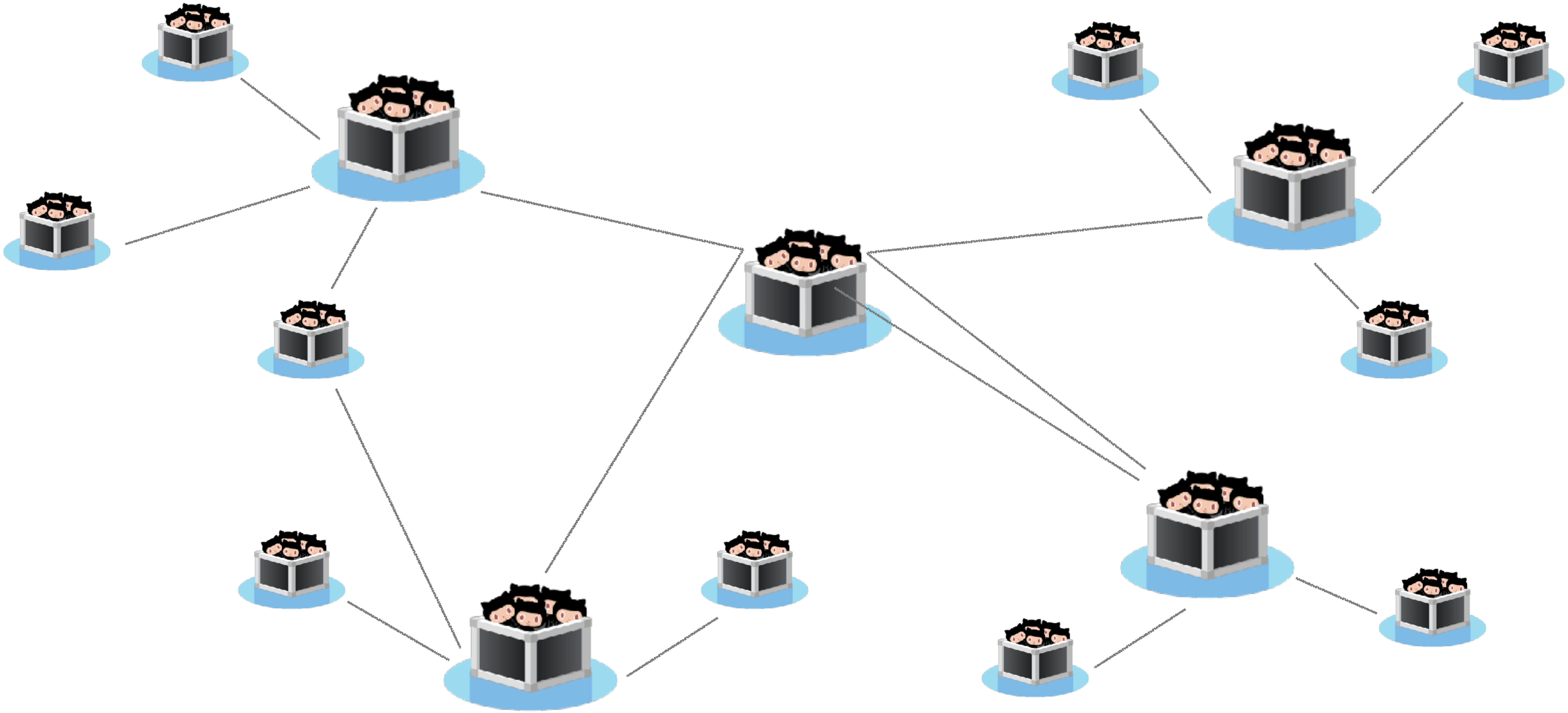
… and the less competition there is with other projects the same people could contribute to

Key labor pool operationalization idea: the collaboration network

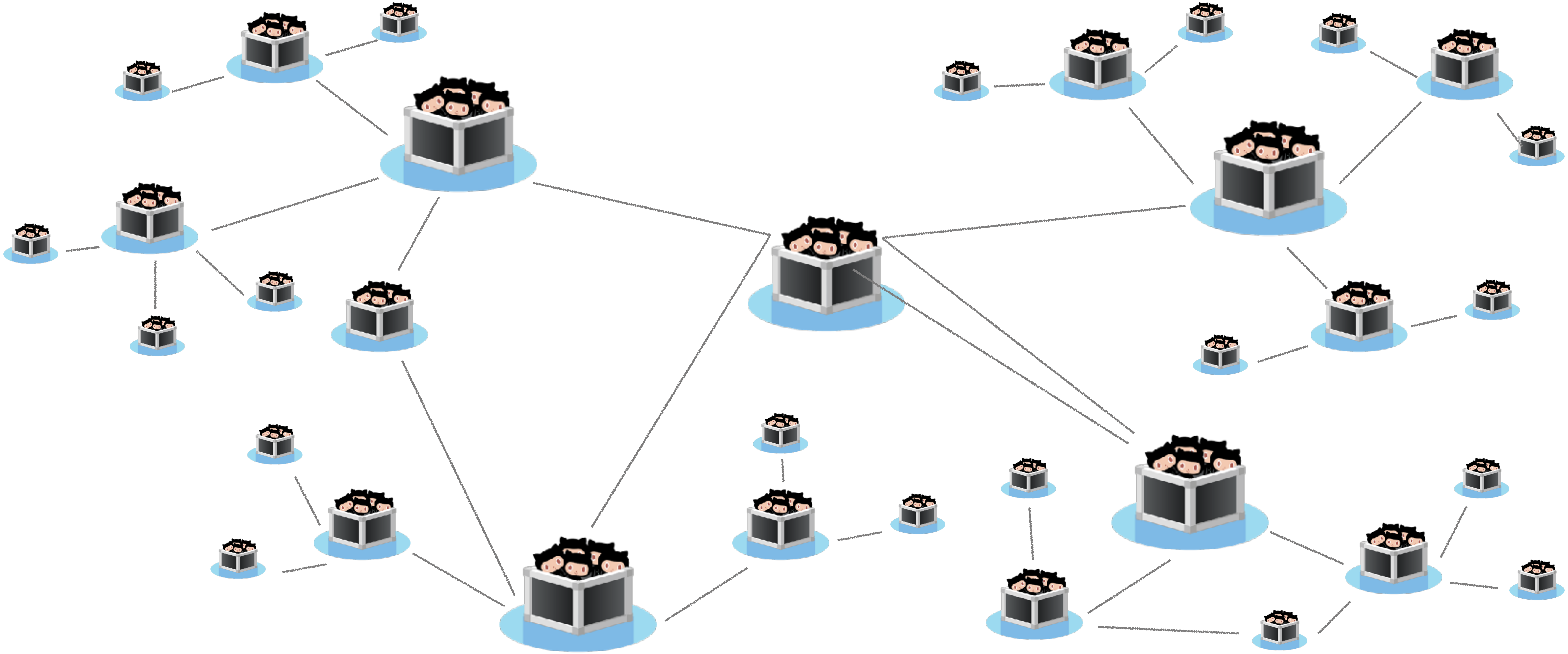
Key labor pool operationalization idea: the collaboration network



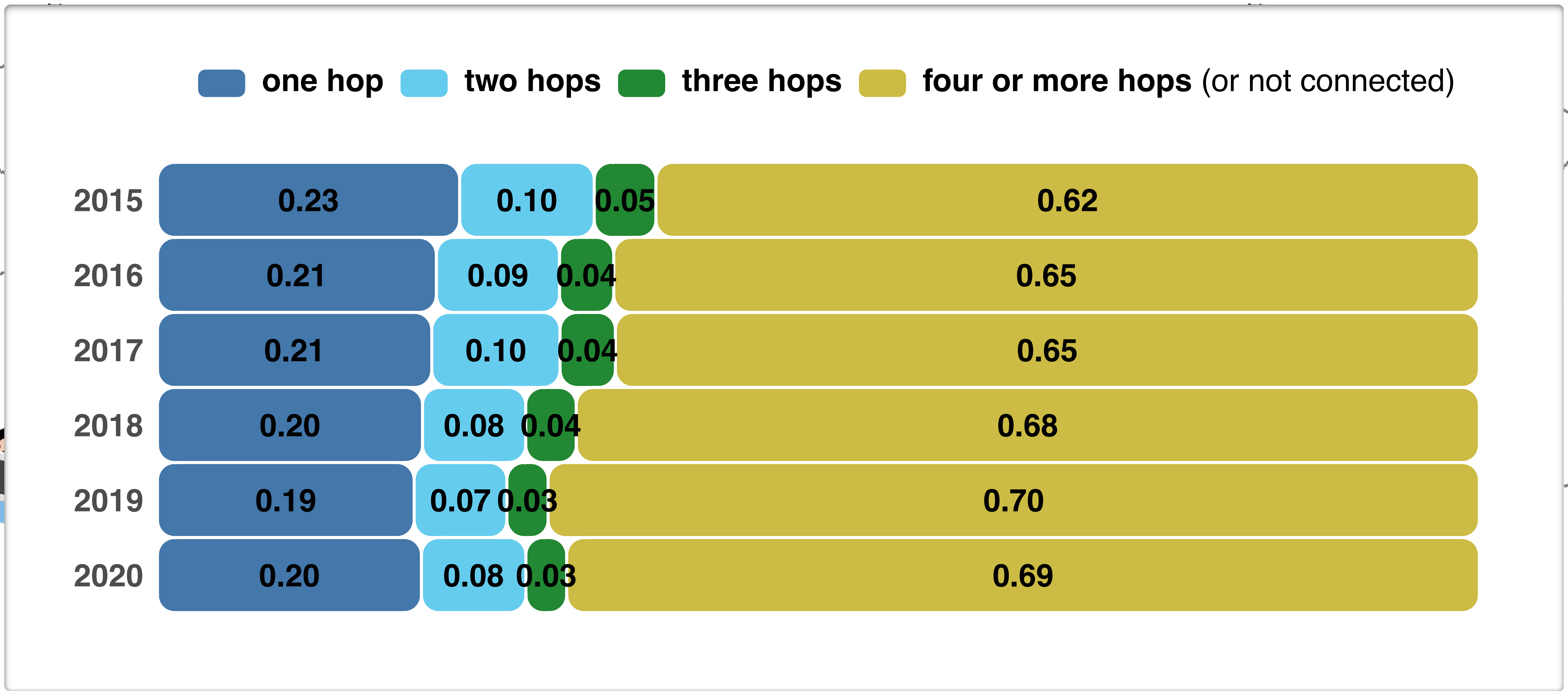
Key labor pool operationalization idea: the collaboration network



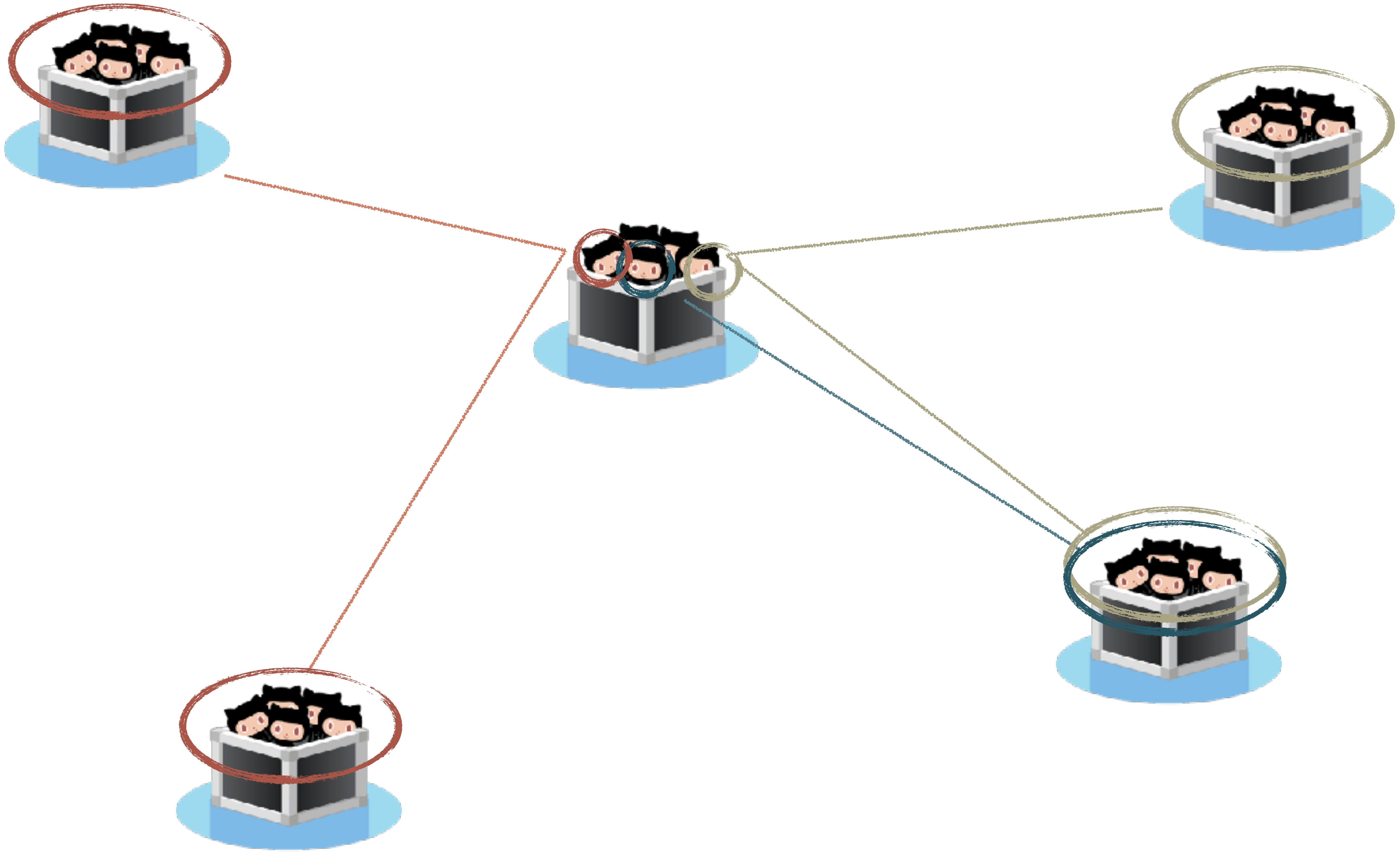
Key labor pool operationalization idea: the collaboration network



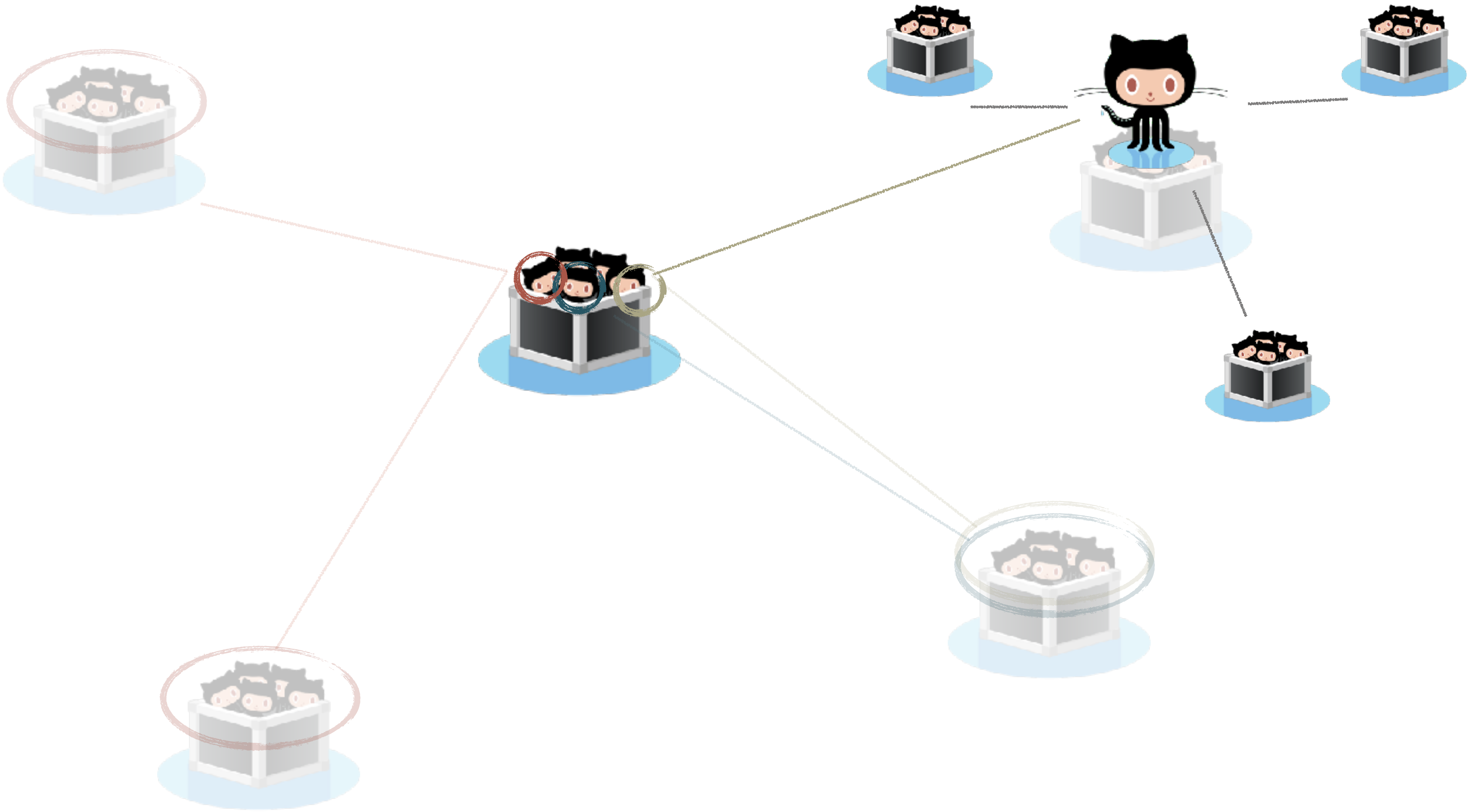
One hop captures 61-65% of everyone identifiable within three hops.



Labor pool operational definition: everyone one hop away in the collaboration network from current project contributors



Labor pool operational definition: everyone one hop away in the collaboration network from current project contributors



For each of these people, we estimate the strength of their social connection to the focal project contributors and their skill match to the focal project, both absolutely and relatively.

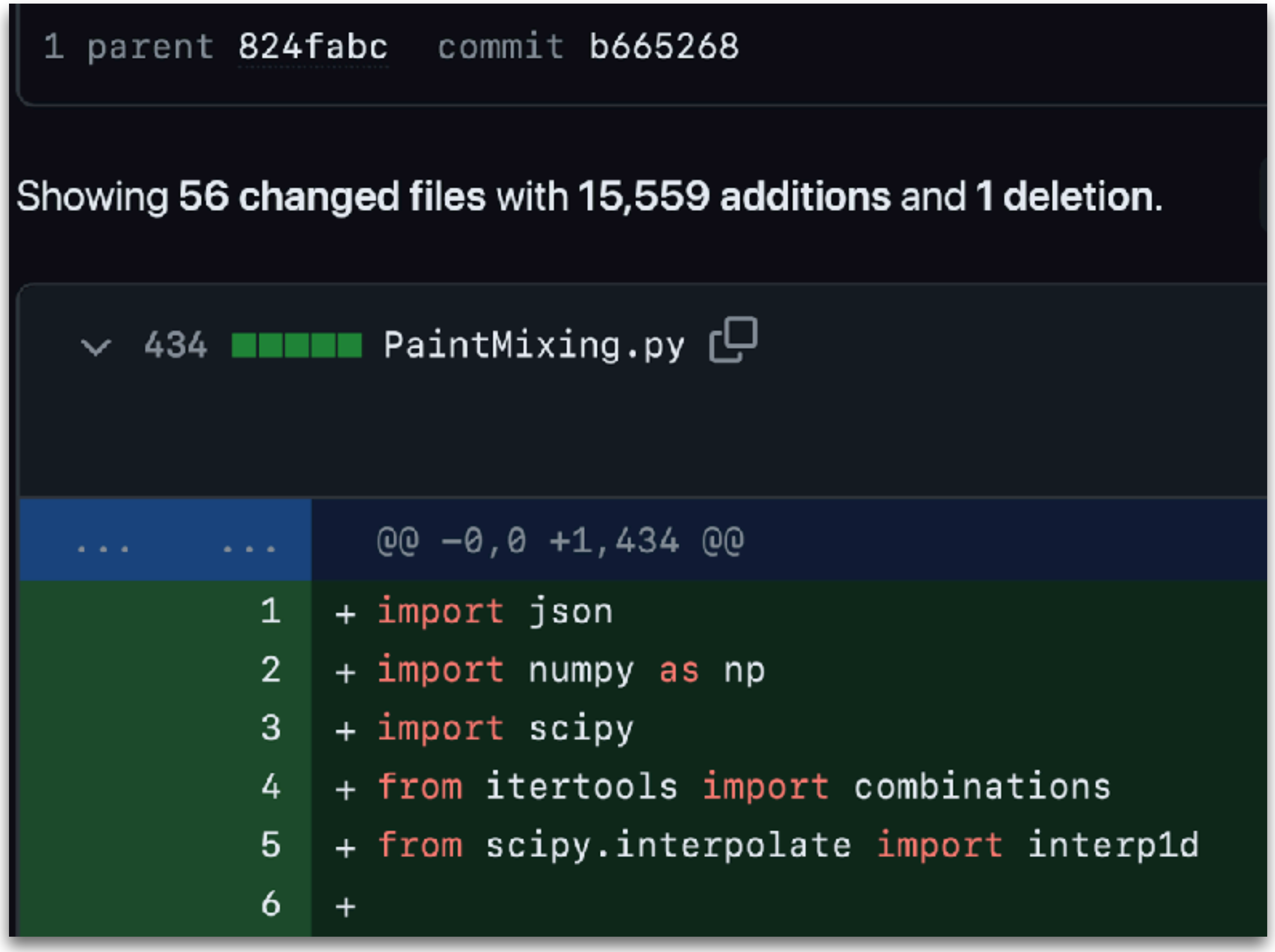
Cosine distance between the developer's and the project's embeddings as a proxy for skill match.

We mine package imports from the commit history to compute technical need / skill (Doc2Vec) embeddings of developers and projects.

Project perspective:

- json
- numpy
- …

across all commits to the project



```
1 parent 824fabc   commit b665268

Showing 56 changed files with 15,559 additions and 1 deletion.

434 ▇▇▇▇▇ PaintMixing.py

...     ...   @@ -0,0 +1,434 @@
           + import json
           + import numpy as np
           + import scipy
           + from itertools import combinations
           + from scipy.interpolate import interp1d
           +
```

Developer perspective:

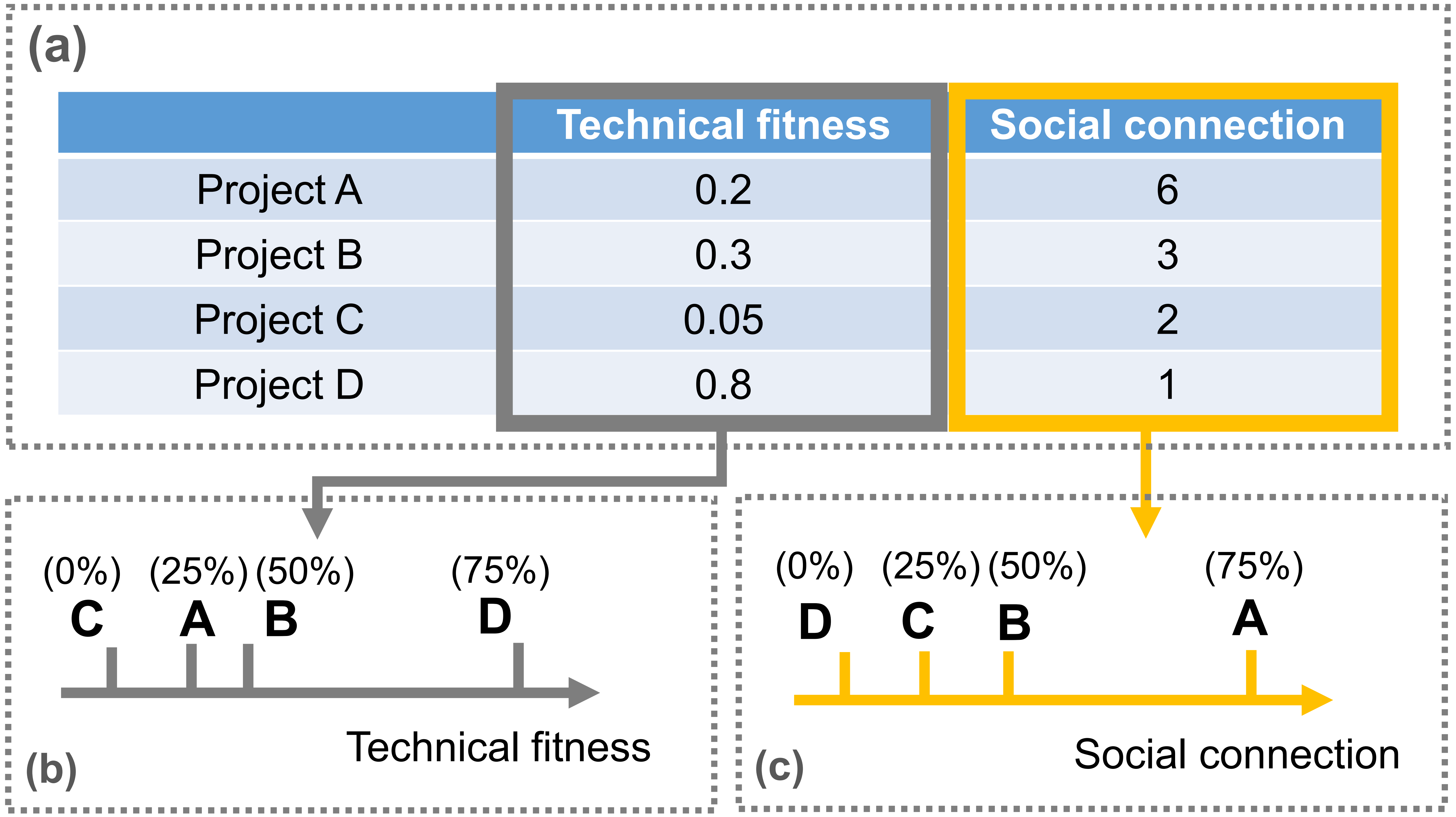
- json
- numpy
- …

across all commits to all projects contributed to

Relative ranking on social connection and technical fit as a proxy for standing with respect to competitors.

Project perspective:

Where do I stand relative to my “competitors”?



	Technical fitness	Social connection
Project A	0.2	6
Project B	0.3	3
Project C	0.05	2
Project D	0.8	1

Developer perspective:

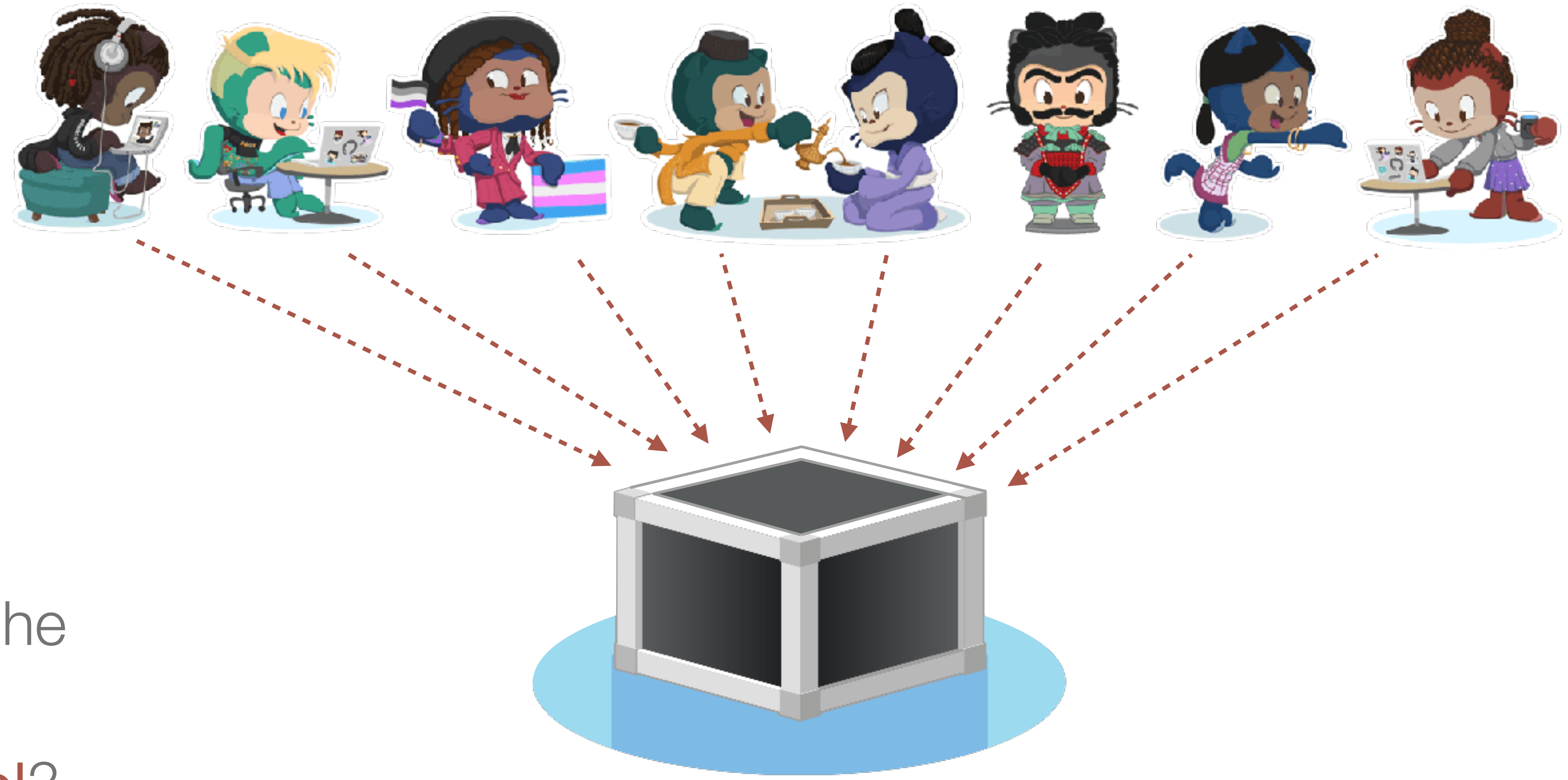
Where does this project stand relative to my other options?

Two-stage regression modeling: individual level + project level

Individual level:

(Logistic regression)

Will this developer contribute to the project next year?



Project level:

(Negative binomial regression)

How many new contributors can the project expect next year?

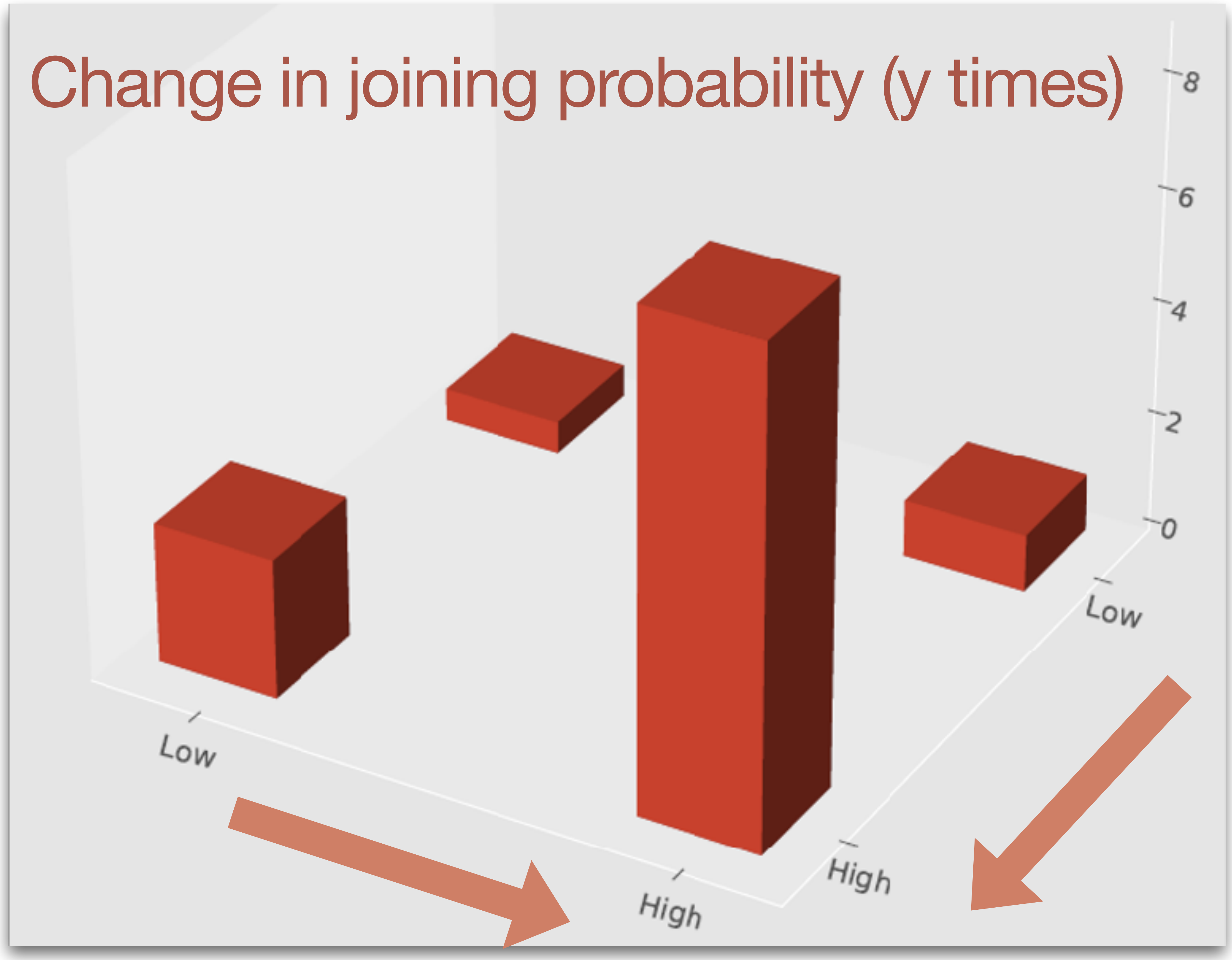
How big is the effective labor pool?

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

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

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

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

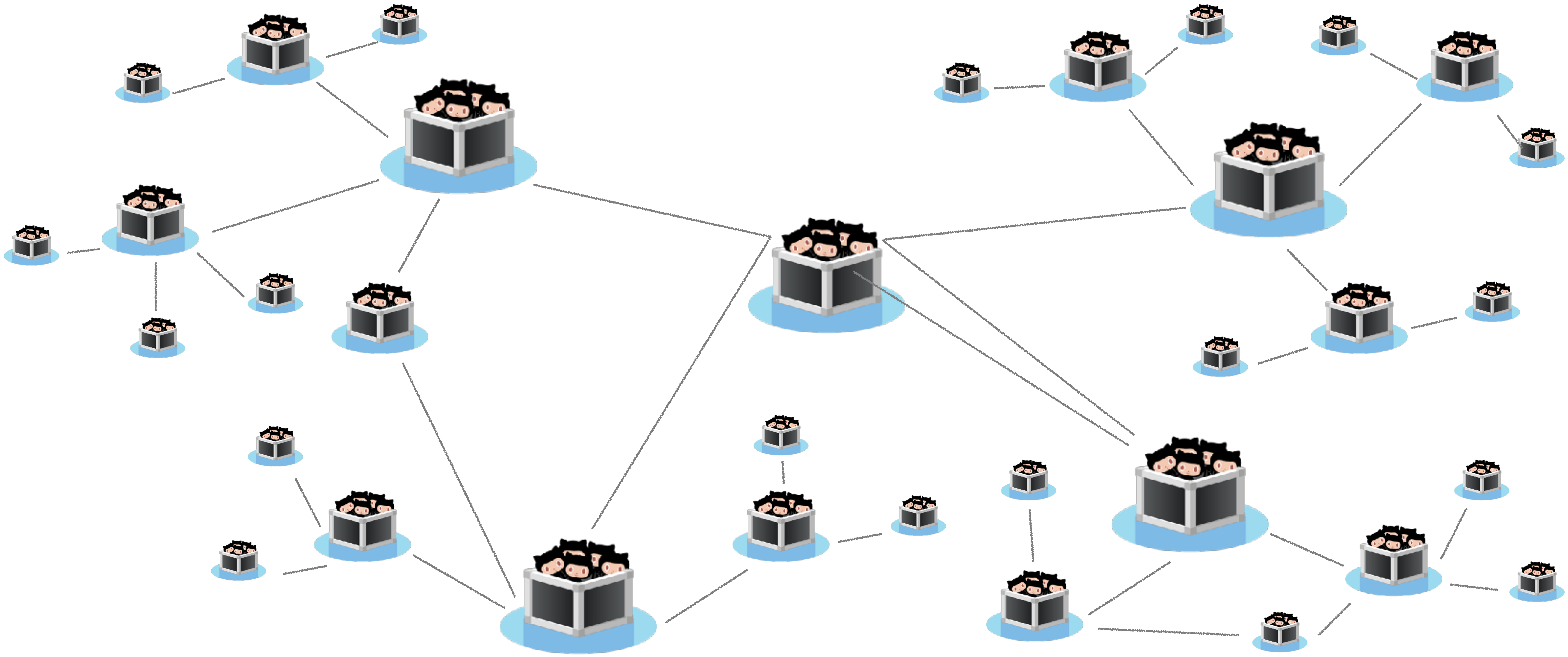


Technical skill match

Social connection strength

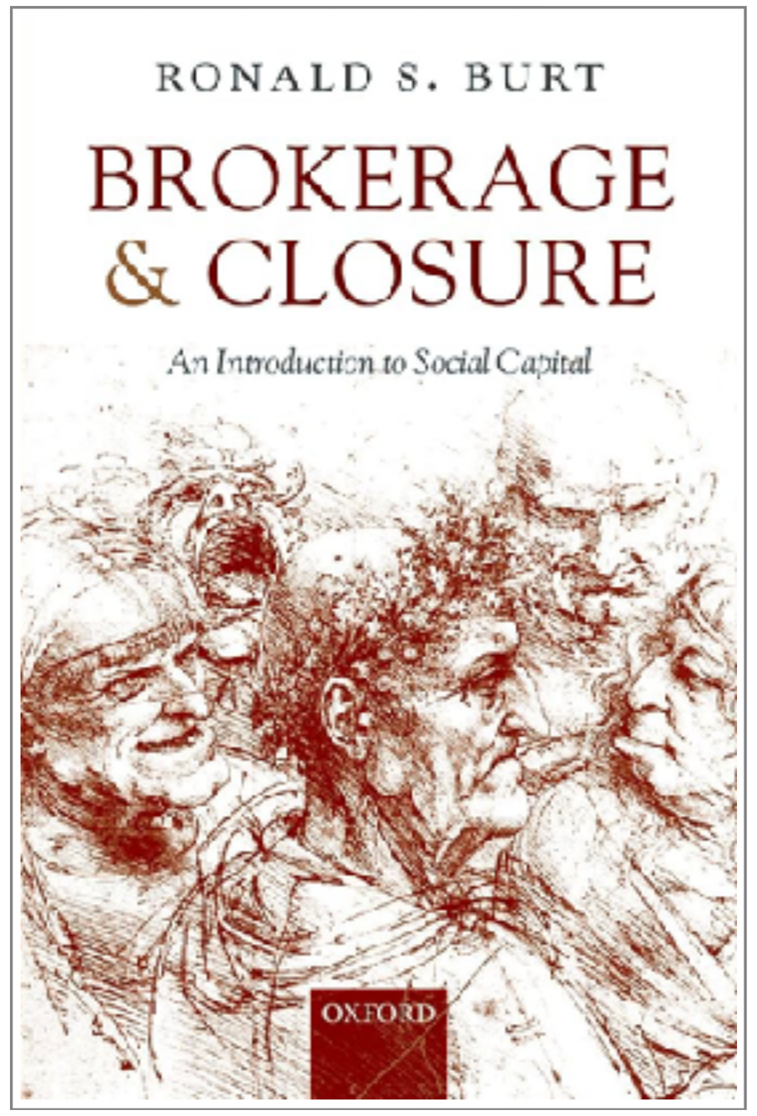
Fine print: Models estimated across 5.78M (contributor, project) pairs. Starting data is all public git repositories in World of Code with 50%+ of their files written in Python, and 10+ commits with import statements.

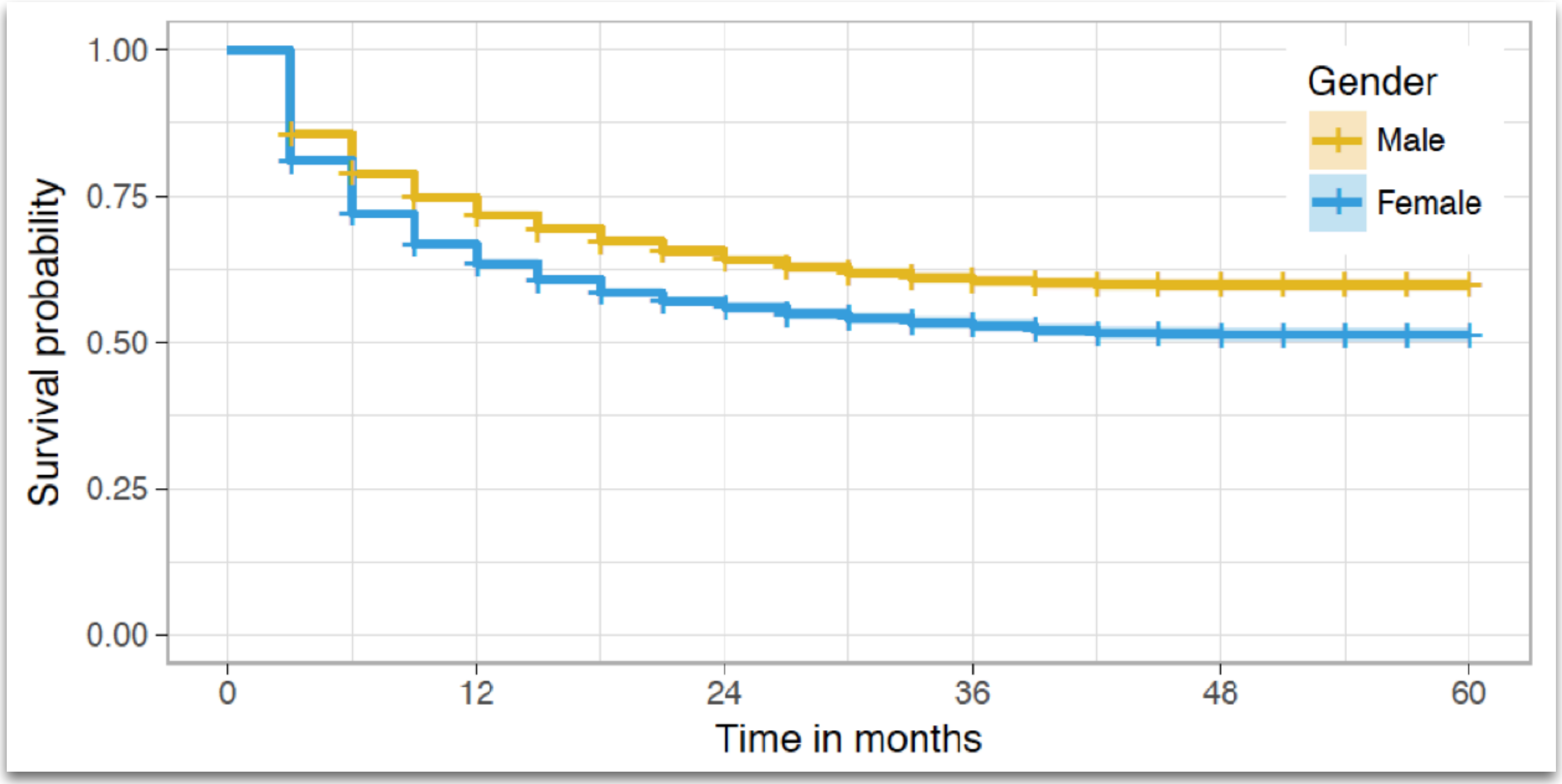
Conclusion: A network-centric perspective reveals interesting ecosystem-level dynamics.



Conclusion: A network-centric perspective reveals interesting ecosystem-level dynamics.

Why do women on GitHub disengage earlier than men?





Qiu, Nolte, Brown, Serebrenik, and Vasilescu. “Going farther together: The impact of social capital on sustained participation in open source.” ICSE 2019 Distinguished Paper Award.

Conclusion: A network-centric perspective reveals interesting ecosystem-level dynamics.

How do tools and practices spread through the network?

12 popular quality assurance tools

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

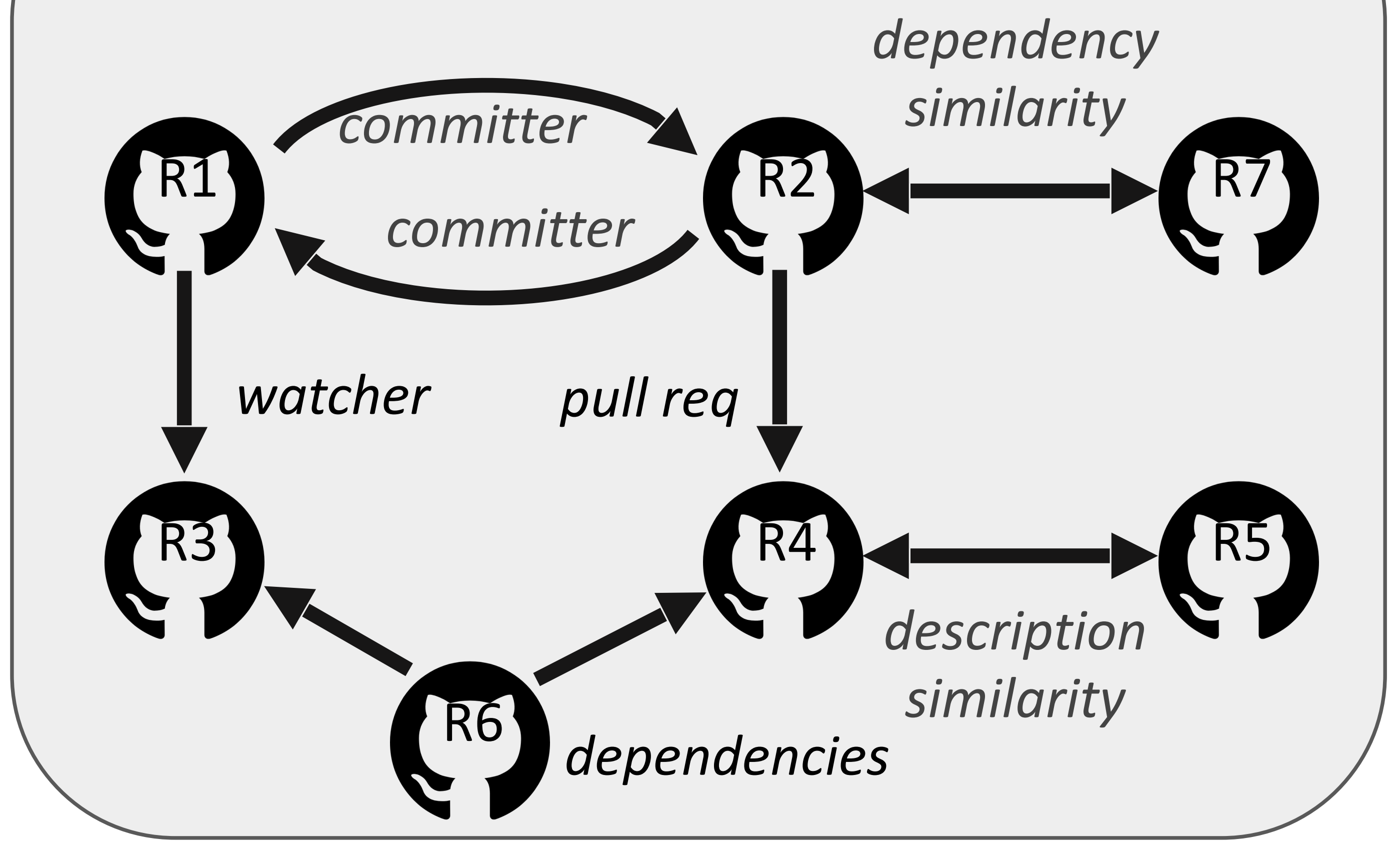
~86,000 npm package repositories

GitHub

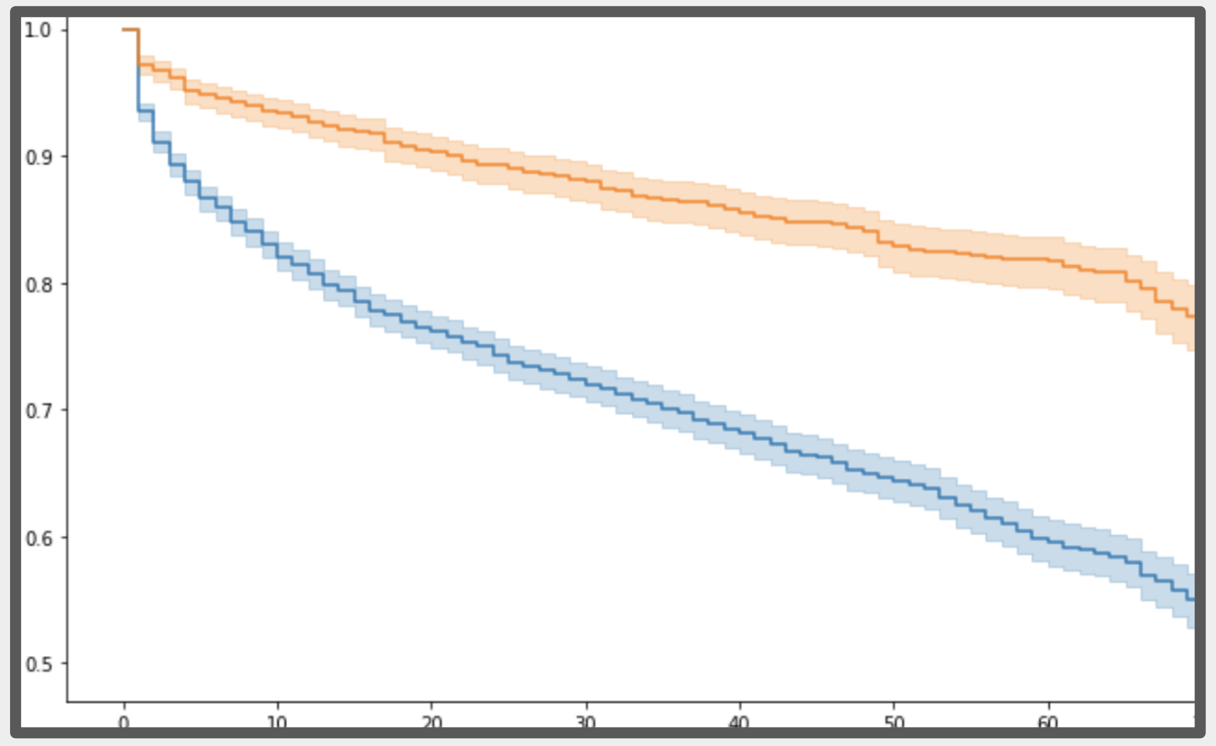
npm

For each tool:

Heterogeneous network



Hazard modeling (Cox regression)



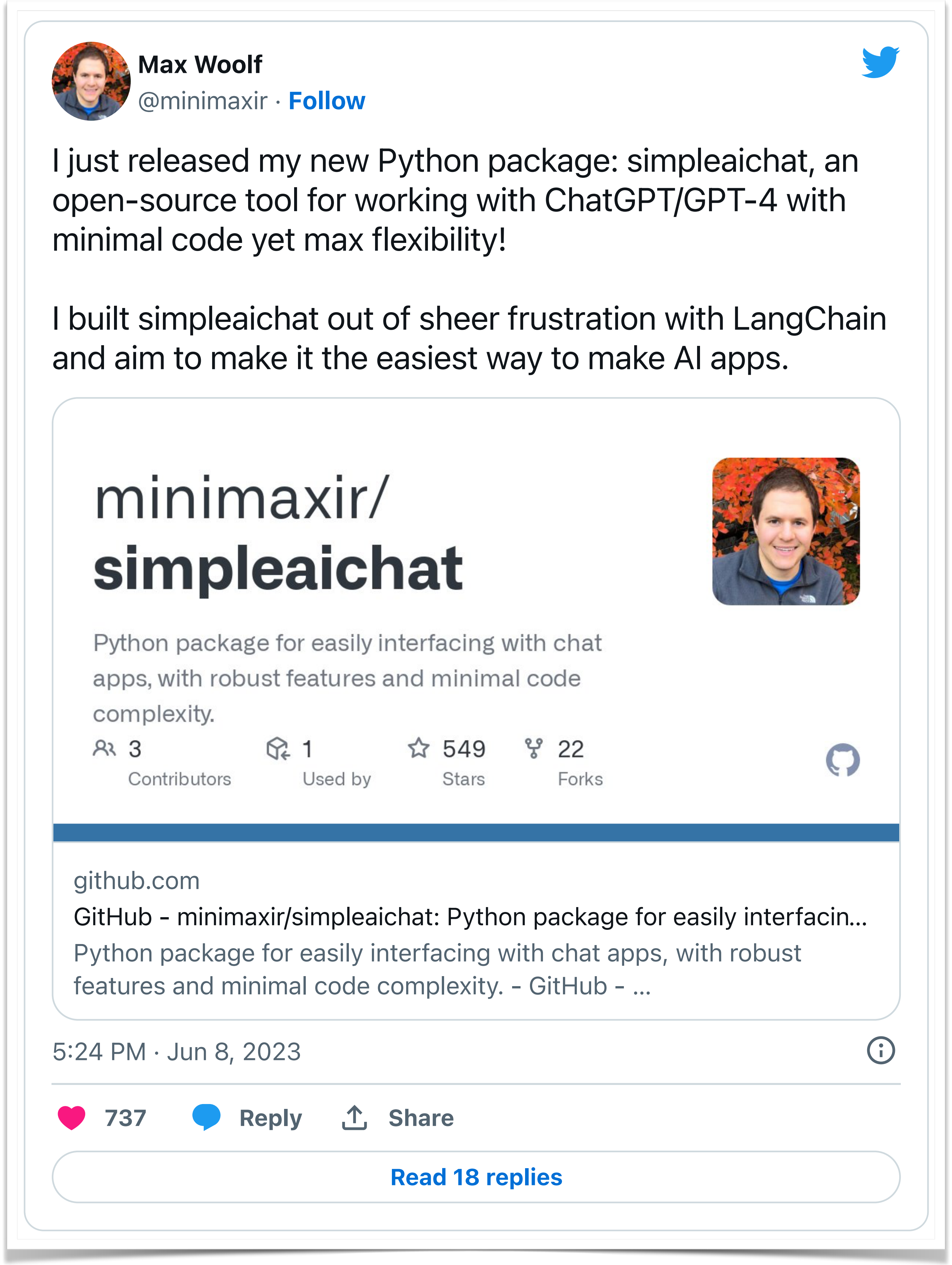
Lamba, Trockman, Armanios, Kästner, Miller, and Vasilescu. “Heard it through the Gitvine: An empirical study of tool diffusion across the npm ecosystem. ESEC/FSE 2020.

Causal Effects of Tweeting

Fang, Lamba, Herbsleb, and Vasilescu. "'This is damn slick!' Estimating the impact of tweets on open source project popularity and new contributors." ICSE 2022. Distinguished Paper Award.

Apple strudel, Beek Cafe, Baden-Baden, Germany 2017

Do tweets cause GitHub stars (and new contributors)?



Max Woolf
@minimaxir · Follow

I just released my new Python package: simpleaichat, an open-source tool for working with ChatGPT/GPT-4 with minimal code yet max flexibility!

I built simpleaichat out of sheer frustration with LangChain and aim to make it the easiest way to make AI apps.

minimaxir/
simpleaichat

Python package for easily interfacing with chat apps, with robust features and minimal code complexity.

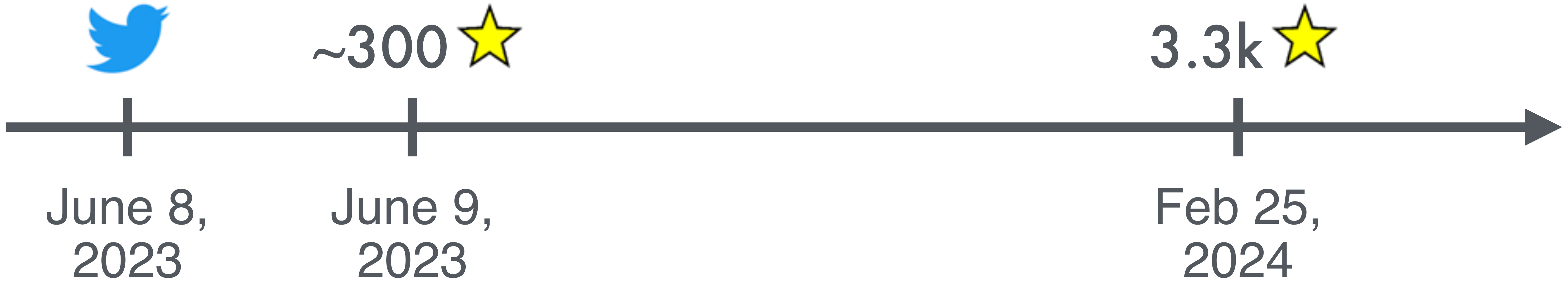
3 Contributors · 1 Used by · 549 Stars · 22 Forks

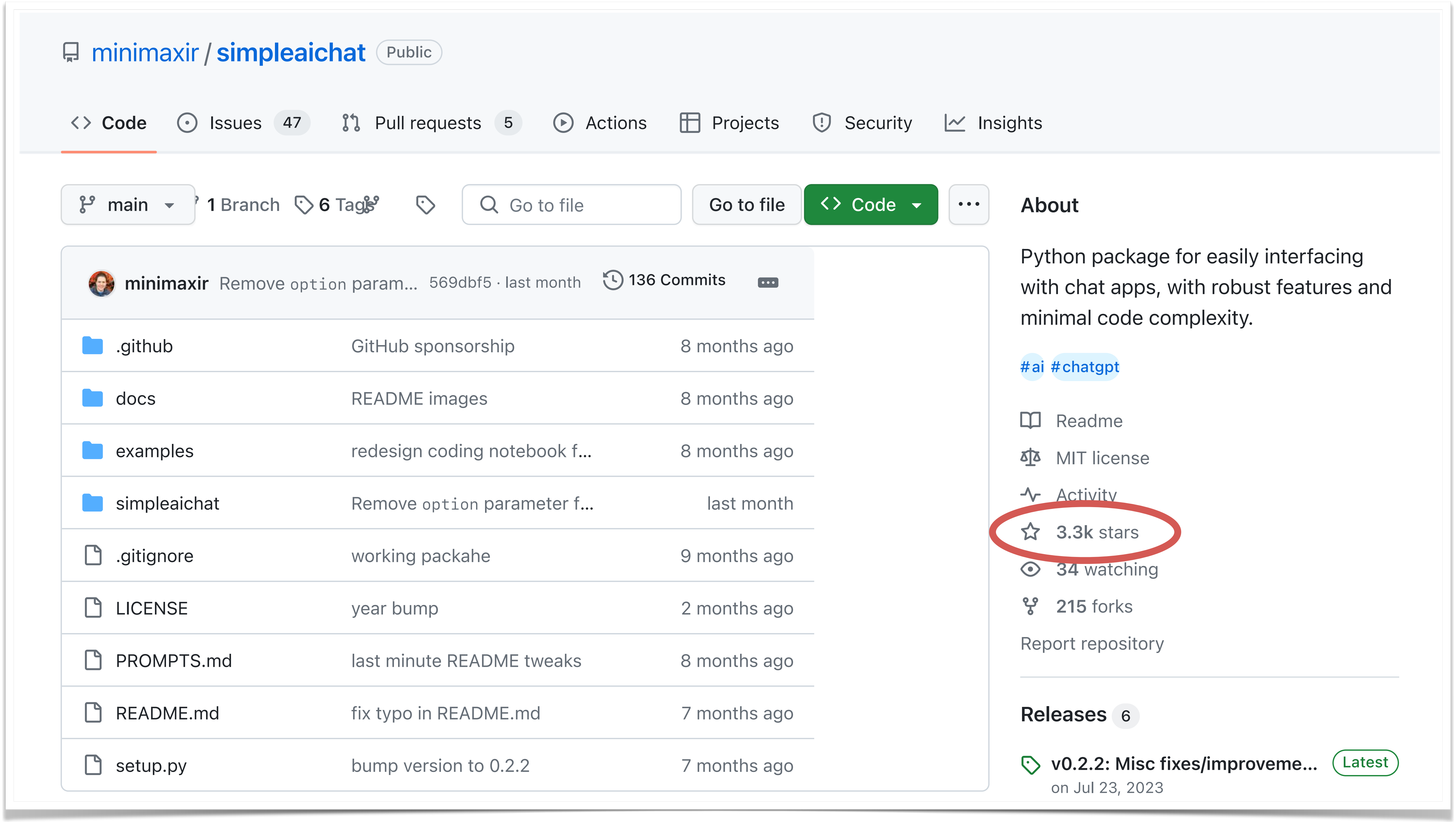
github.com
GitHub - minimaxir/simpleaichat: Python package for easily interfacin...
Python package for easily interfacing with chat apps, with robust features and minimal code complexity. - GitHub - ...

5:24 PM · Jun 8, 2023

737 · Reply · Share

Read 18 replies





minimaxir / **simpleaichat** Public

Code · Issues 47 · Pull requests 5 · Actions · Projects · Security · Insights

main · 1 Branch · 6 Tags · Go to file · Go to file · Code

minimaxir Remove `option` param... 569dbf5 · last month · 136 Commits

.github	GitHub sponsorship	8 months ago
docs	README images	8 months ago
examples	redesign coding notebook f...	8 months ago
simpleaichat	Remove `option` parameter f...	last month
.gitignore	working packahe	9 months ago
LICENSE	year bump	2 months ago
PROMPTS.md	last minute README tweaks	8 months ago
README.md	fix typo in README.md	7 months ago
setup.py	bump version to 0.2.2	7 months ago

About

Python package for easily interfacing with chat apps, with robust features and minimal code complexity.

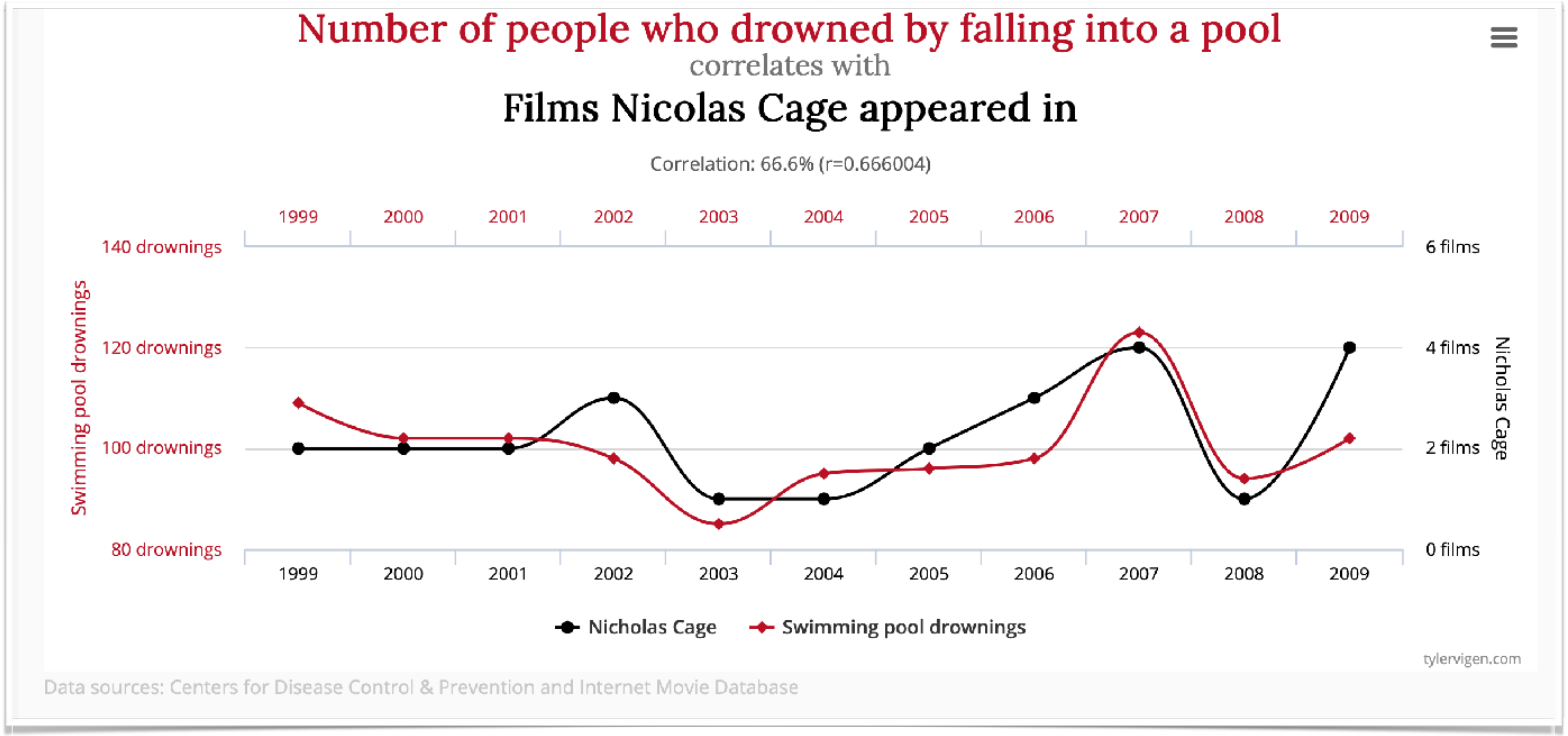
#ai #chatgpt

Readme
MIT license
Activity
3.3k stars
34 watching
215 forks
Report repository

Releases 6

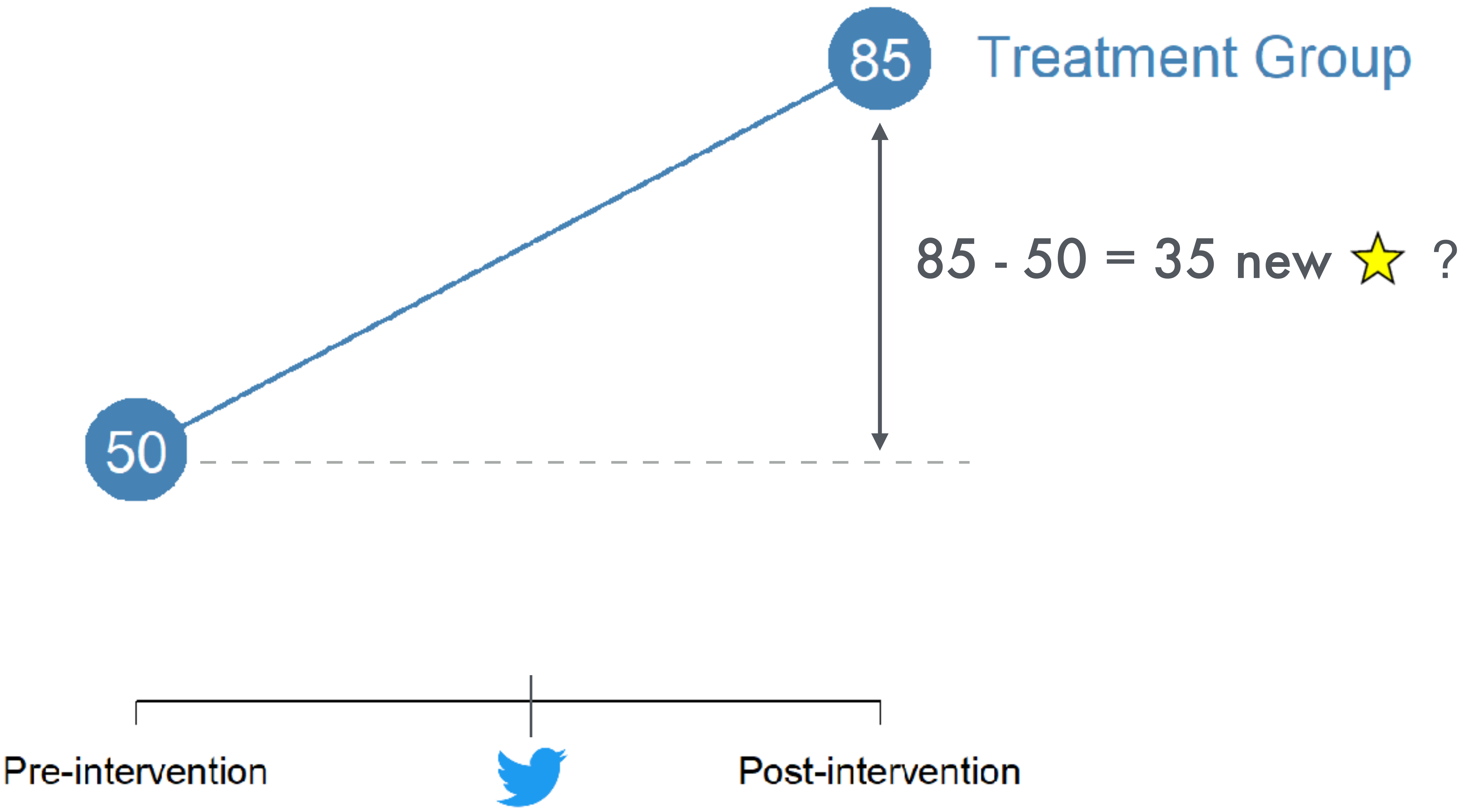
v0.2.2: Misc fixes/improveme... Latest
on Jul 23, 2023

Do Nicolas Cage movies cause drowning?

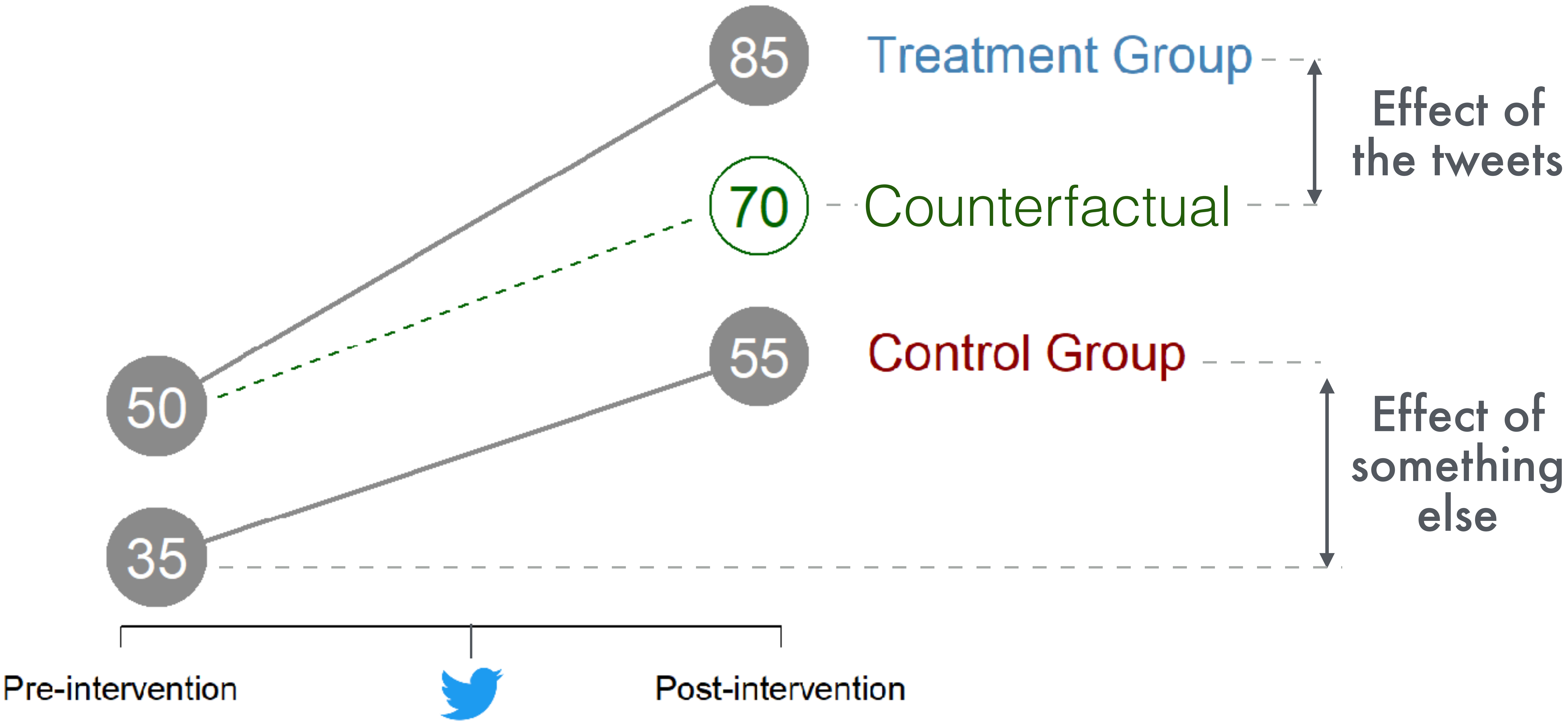


http://www.tylervigen.com/spurious-correlations

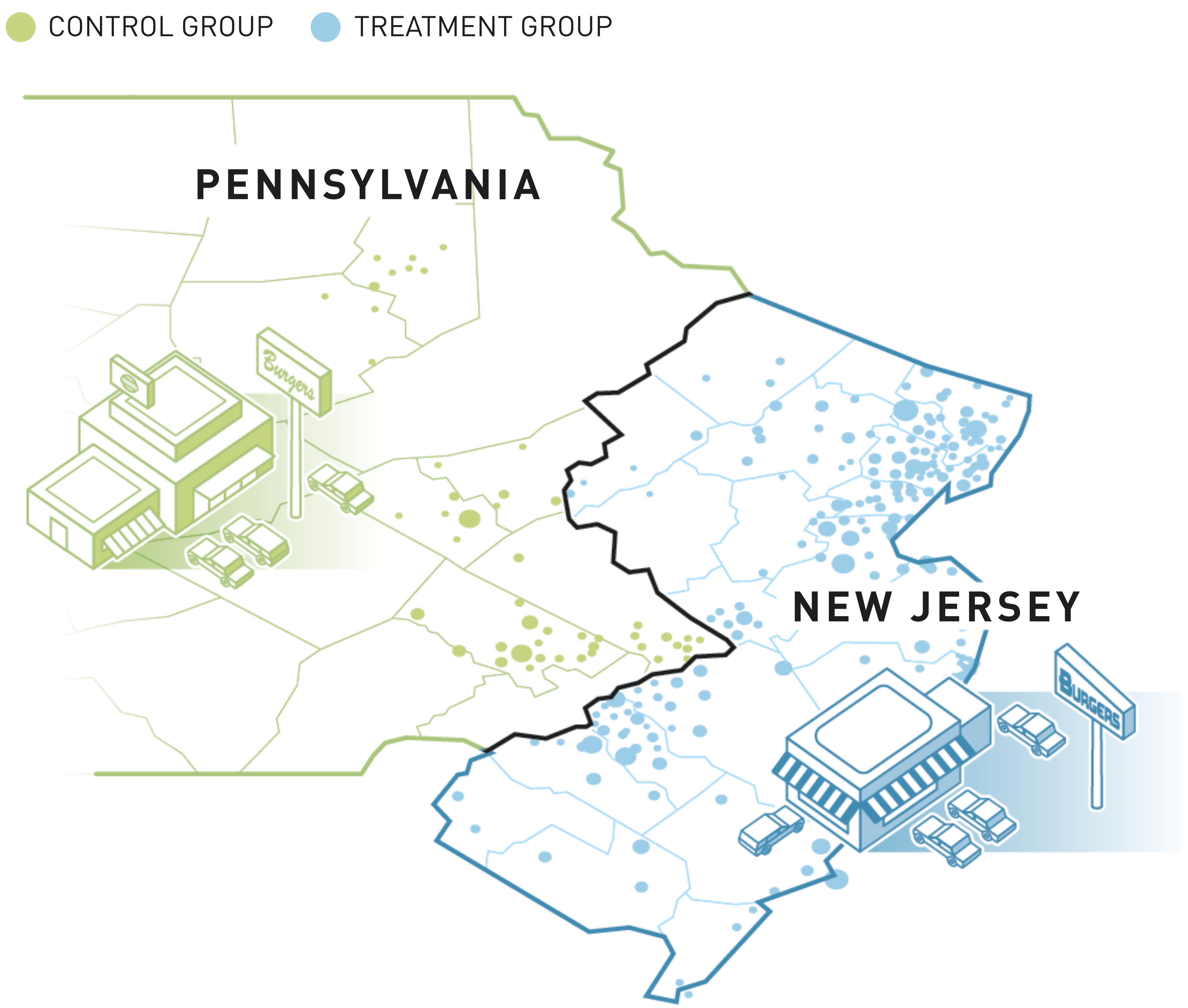
Idea: Measure how much a group mean changes before and after an intervention

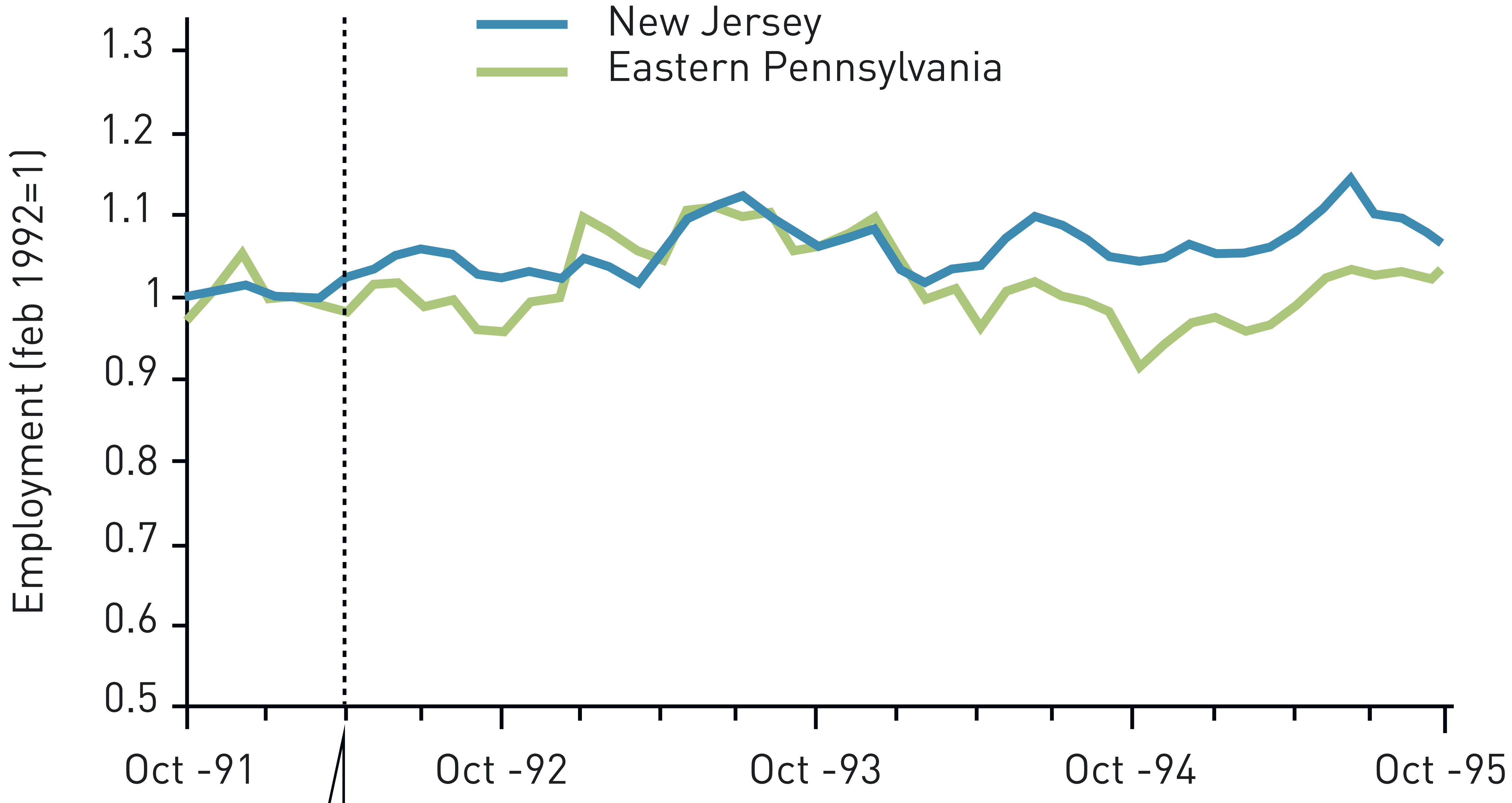


Better idea: Compare that change to the change in an appropriate control group



Card and Krueger (1993) natural experiment to study how increasing the minimum wage affects employment.



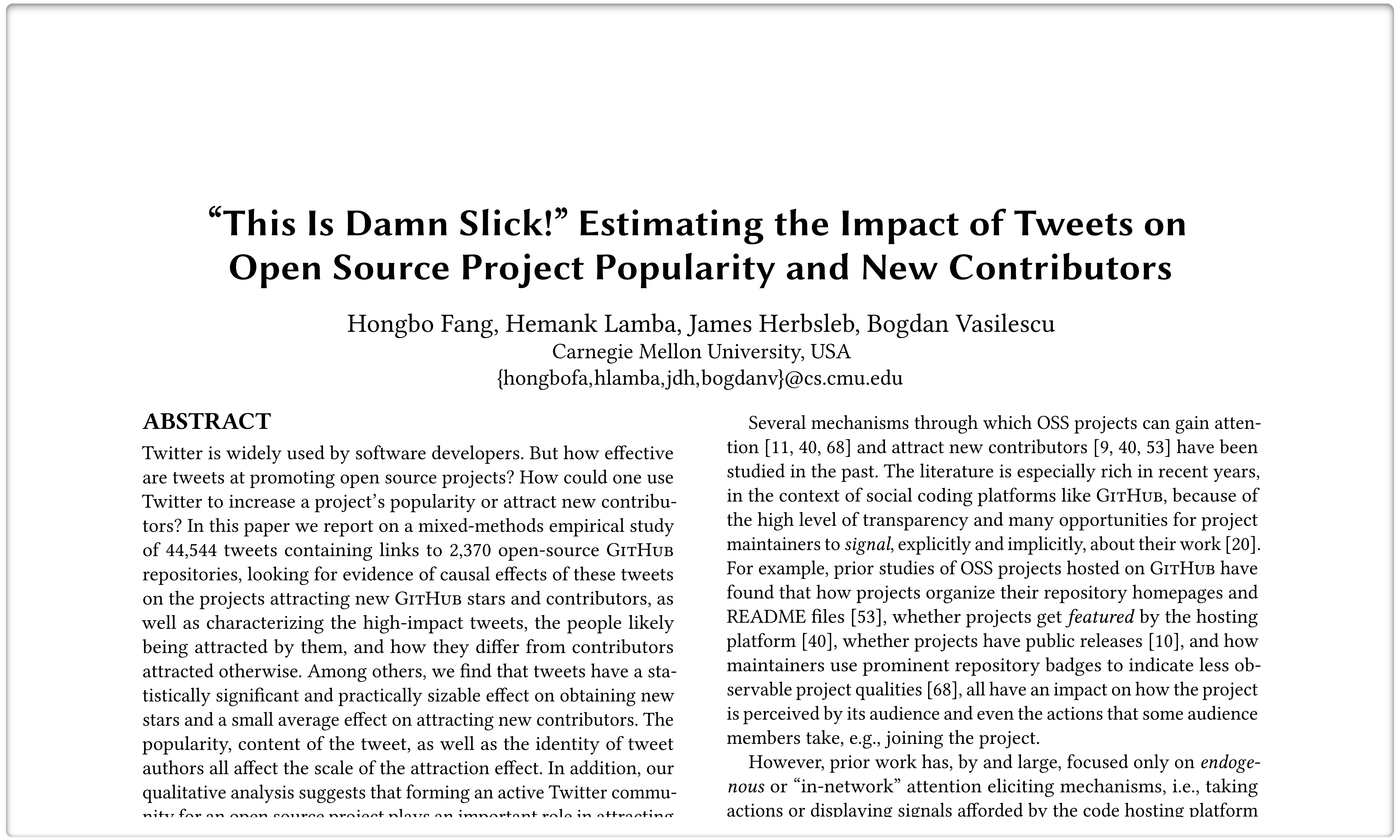


1 April 1992: The hourly minimum wage in New Jersey was increased from 4.25 dollars to 5.05 dollars. Despite this, employment in New Jersey was not affected.

NOBEL PRIZE ECONOMICS

https://www.nobelprize.org/uploads/2021/10/popular-economicsciencesprize2021-2.pdf

Aside: Are we 20 years behind on empirical methods in SE?



“This Is Damn Slick!” Estimating the Impact of Tweets on Open Source Project Popularity and New Contributors

Hongbo Fang, Hemank Lamba, James Herbsleb, Bogdan Vasilescu
Carnegie Mellon University, USA
{hongbofa,hlamba,jdh,bogdanv}@cs.cmu.edu

ABSTRACT

Twitter is widely used by software developers. But how effective are tweets at promoting open source projects? How could one use Twitter to increase a project’s popularity or attract new contributors? In this paper we report on a mixed-methods empirical study of 44,544 tweets containing links to 2,370 open-source GitHub repositories, looking for evidence of causal effects of these tweets on the projects attracting new GitHub stars and contributors, as well as characterizing the high-impact tweets, the people likely being attracted by them, and how they differ from contributors attracted otherwise. Among others, we find that tweets have a statistically significant and practically sizable effect on obtaining new stars and a small average effect on attracting new contributors. The popularity, content of the tweet, as well as the identity of tweet authors all affect the scale of the attraction effect. In addition, our qualitative analysis suggests that forming an active Twitter community for an open source project plays an important role in attracting

Several mechanisms through which OSS projects can gain attention [11, 40, 68] and attract new contributors [9, 40, 53] have been studied in the past. The literature is especially rich in recent years, in the context of social coding platforms like GitHub, because of the high level of transparency and many opportunities for project maintainers to *signal*, explicitly and implicitly, about their work [20]. For example, prior studies of OSS projects hosted on GitHub have found that how projects organize their repository homepages and README files [53], whether projects get *featured* by the hosting platform [40], whether projects have public releases [10], and how maintainers use prominent repository badges to indicate less observable project qualities [68], all have an impact on how the project is perceived by its audience and even the actions that some audience members take, e.g., joining the project.

However, prior work has, by and large, focused only on *endogenous* or “in-network” attention eliciting mechanisms, i.e., taking actions or displaying signals afforded by the code hosting platform

Difference in differences (ICSE 2022)

2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)

Do Developers Discover New Tools On The Toilet?

Emerson Murphy-Hill *Google, LLC* emersonm@google.com	Edward K. Smith* *Bloomberg* tedks@riseup.net	Caitlin Sadowski *Google, LLC* supertri@google.com	Ciera Jaspan *Google, LLC* ciera@google.com	Collin Winter* *Waymo* collinwinter@waymo.com
Matthew Jorde *Google, LLC* majorde@google.com	Andrea Knight *Google, LLC* aknight@google.com	Andrew Trenk *Google, LLC* atrenk@google.com	Steve Gross *Google, LLC* stevegross@google.com	

***Abstract*—Maintaining awareness of useful tools is a substantial challenge for developers. Physical newsletters are a simple technique to inform developers about tools. In this paper, we evaluate such a technique, called Testing on the Toilet, by performing a mixed-methods case study. We first quantitatively evaluate how effective this technique is by applying statistical causal inference over six years of data about tools used by thousands of developers. We then qualitatively contextualize these results by interviewing and surveying 382 developers, from authors to editors to readers. We found that the technique was generally effective at increasing software development tool use, although the increase**

Testing on the Toilet Presents... Healthy Code on the Commode

Episode 284
April 30 2013

Automatic formatting for C++

by Daniel Jasper in Munich

Are you tired of hitting space and backspace more often then anything else while coding? Are you **annoyed by fighting over parameter and comment alignment in code reviews?**

Consistent formatting allows readers to quickly scan and interpret code, dedicating their attention to what the code does and how it works. Without this consistency, effort is wasted parsing the wide variety of personal styles code might follow. However, **keeping your code formatting nice and shiny is not a good task for humans**. Luckily, we now have clang-format, which can do this tedious task for you.

Clang-format produces both readable and Google style-compliant code:

```
$ cat file.cc
int a;// clang-format can ..
int bbb;                  // .. align trailing comments.
```

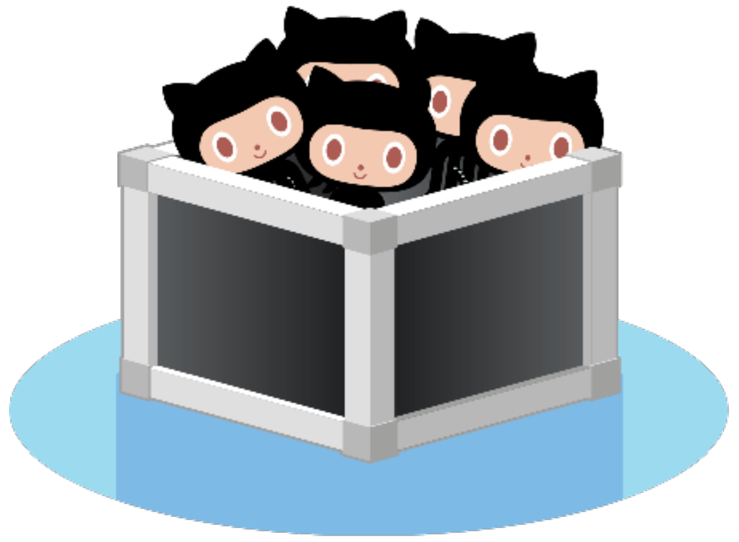
CausalImpact (ICSE 2019)

Yes! Tweets cause stars and new contributors.

SPOILER ALERT

+7%

(+1.2 stars every tweet burst)



+2%

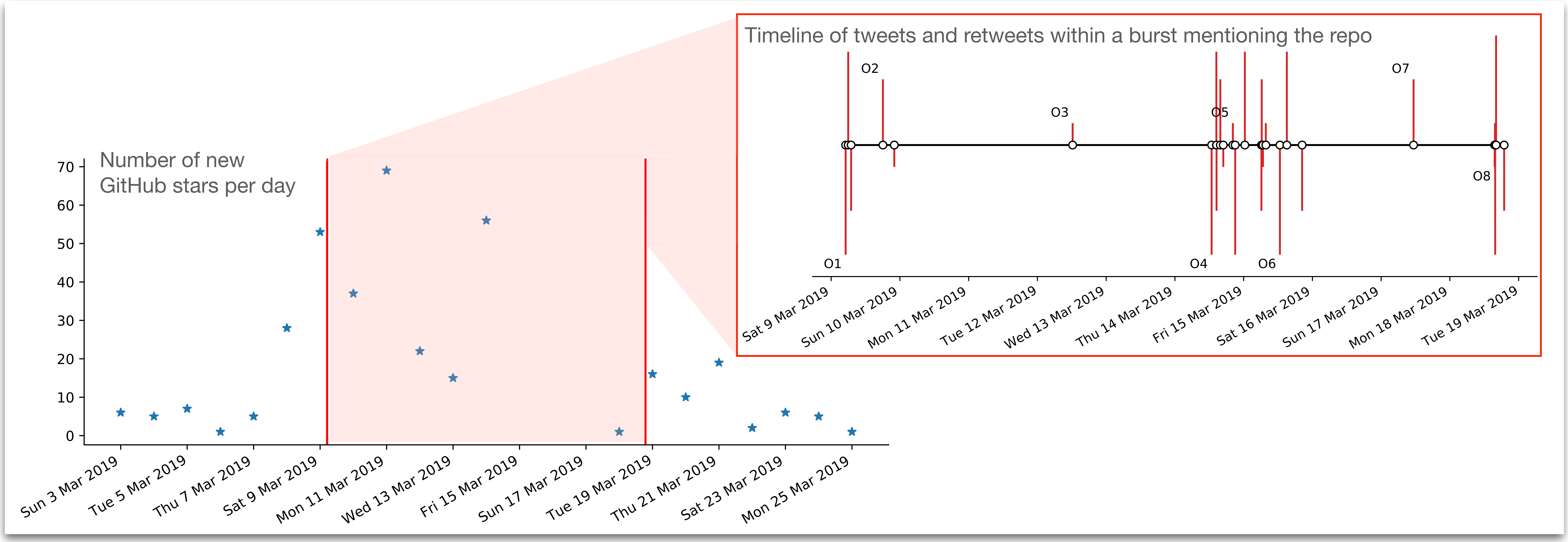
(+1 new contributor every 250 tweet bursts)

Fine print: Models estimated across 2,370 GitHub projects mentioned in 44,544 tweets.

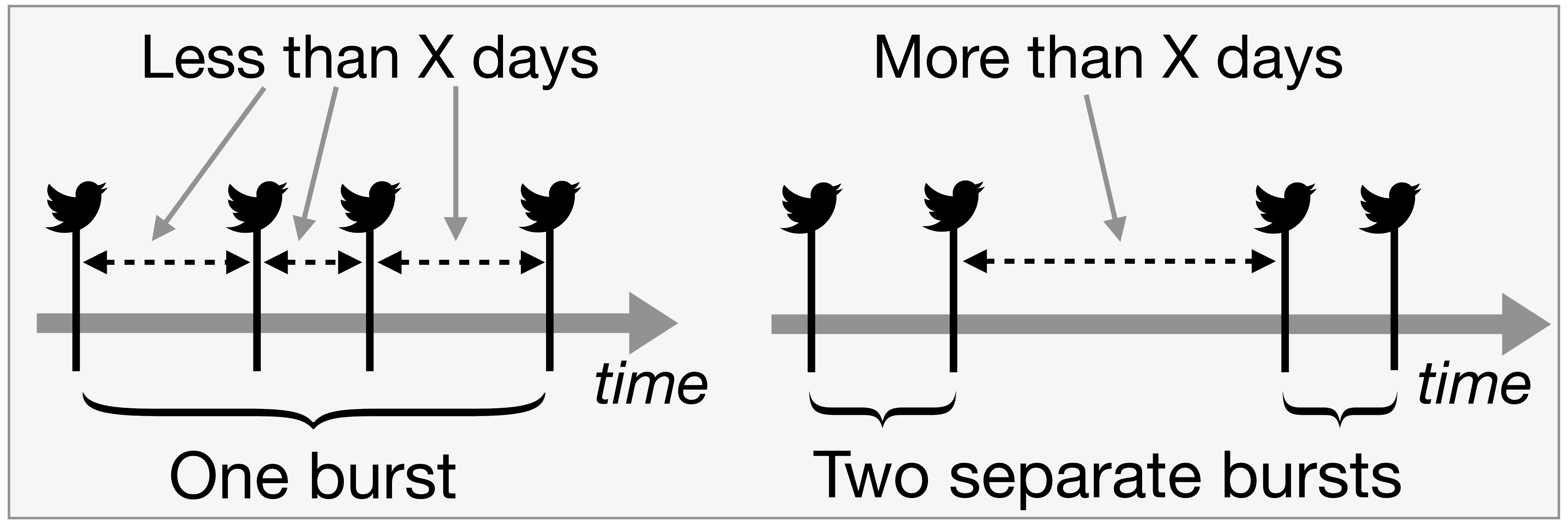
How to actually measure these effects?

Challenge: (Usually) More than one tweet. What should count?

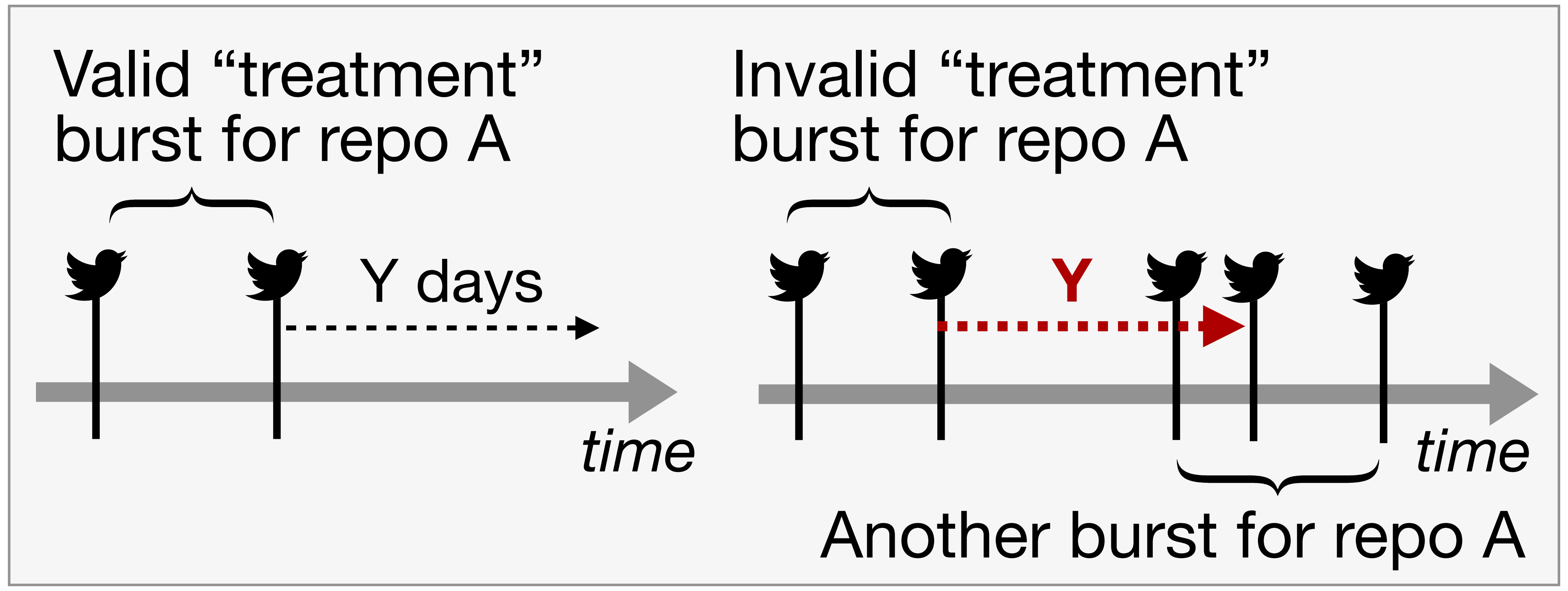
Many heuristics to group tweets into “bursts.” Manual validation + sensitivity analysis.



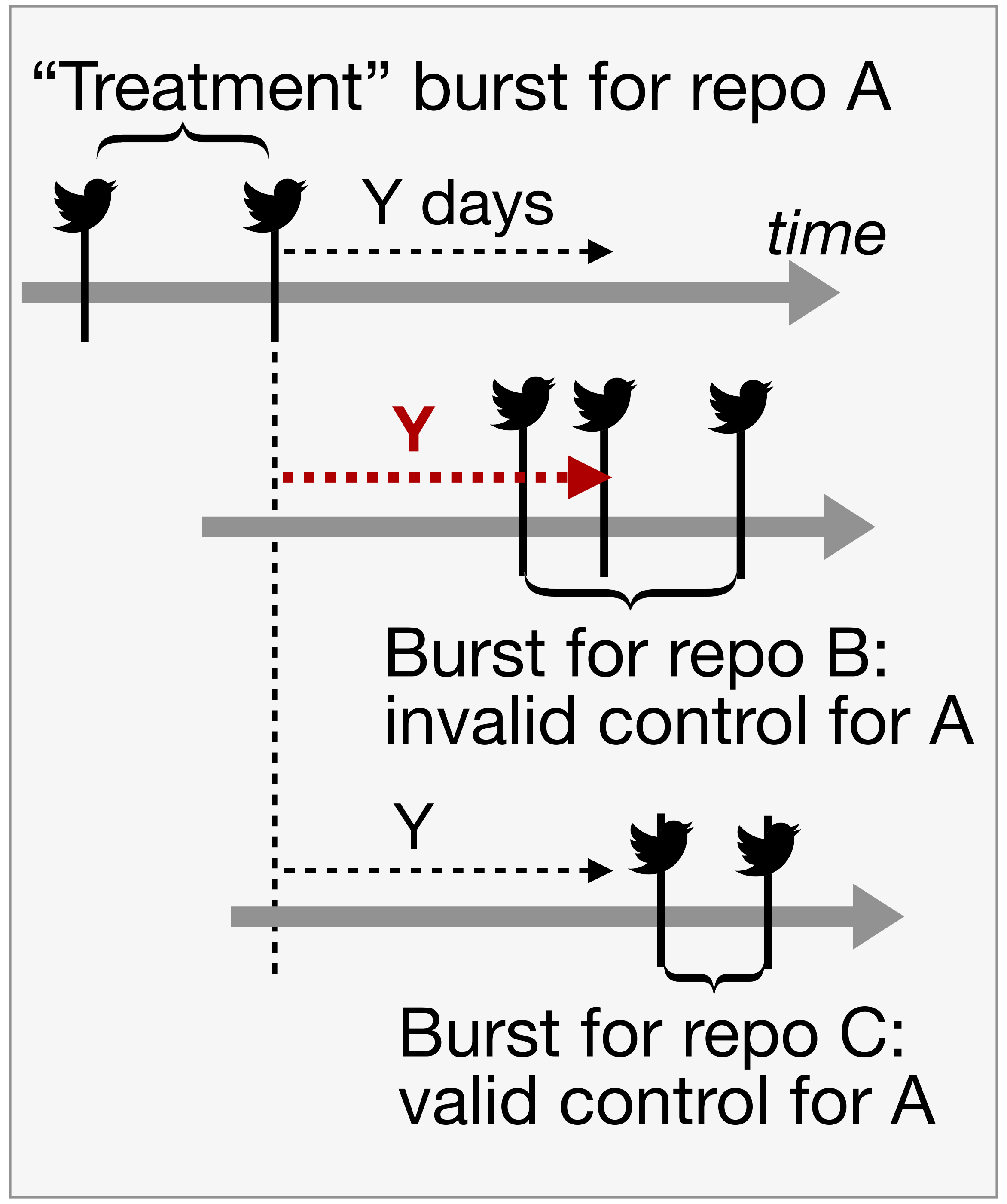
Challenge: (Usually) More than one tweet. What should count?



Tweets mentioning the same project within X days of each other are considered part of the same burst.


Valid "treatment" burst for repo A
Invalid "treatment" burst for repo A
Y days
Y
time
time
Another burst for repo A

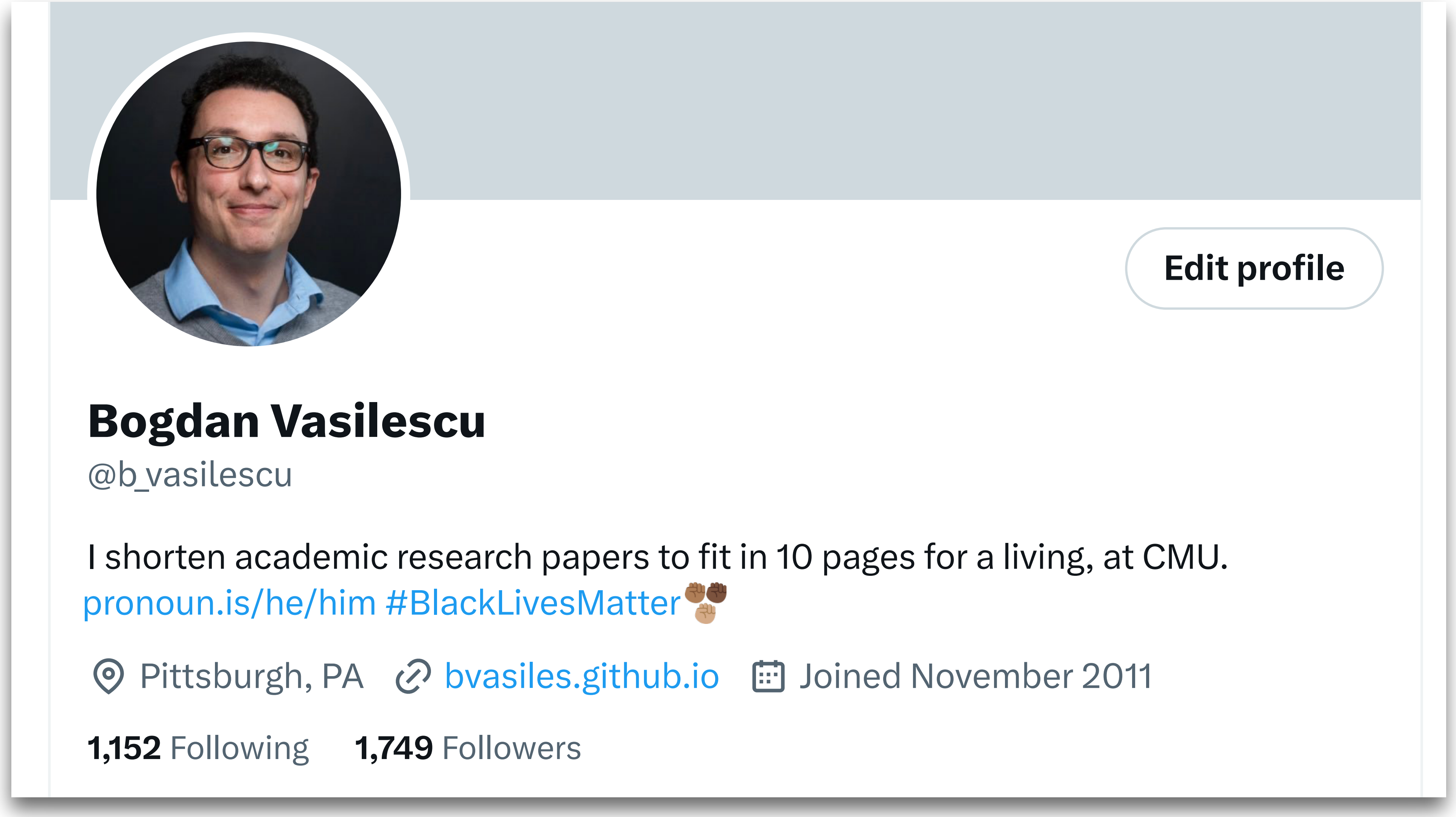
Two bursts mentioning the same project must be at least Y days apart.


"Treatment" burst for repo A
Y days
time
Y
Burst for repo B: invalid control for A
Y
Burst for repo C: valid control for A

Control group repositories must not have experienced any bursts of their own at least Y days after the end of the corresponding treatment group repository burst.

Challenge: Merging identities

Many heuristics, manually validated, to cross-link users between the two platforms.



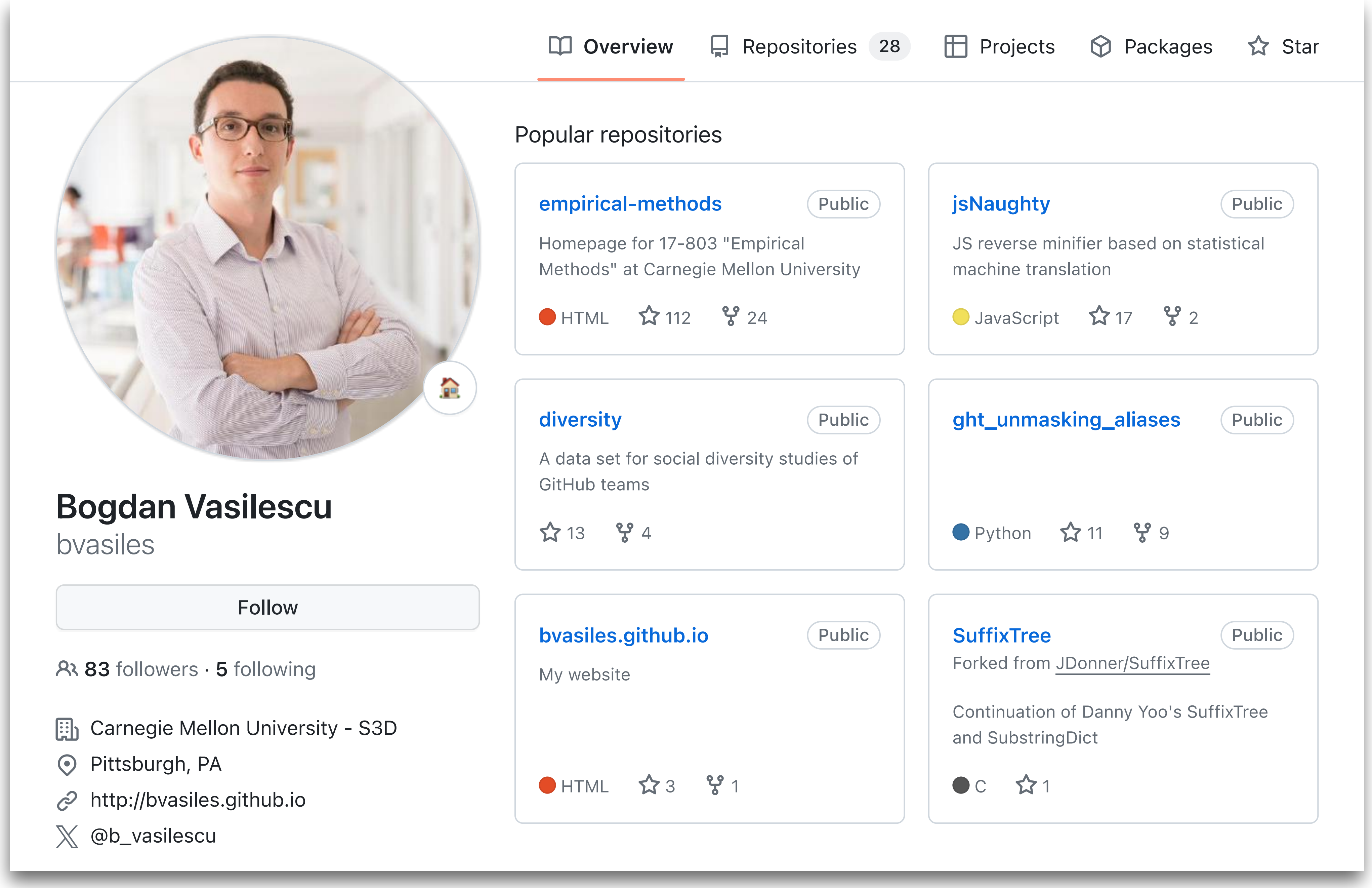
Edit profile

Bogdan Vasilescu
@b_vasilescu

I shorten academic research papers to fit in 10 pages for a living, at CMU. pronoun.is/he/him #BlackLivesMatter

Pittsburgh, PA · bvasiles.github.io · Joined November 2011

1,152 Following **1,749** Followers



Overview · Repositories 28 · Projects · Packages · Star

Bogdan Vasilescu
bvasiles

Follow

83 followers · 5 following

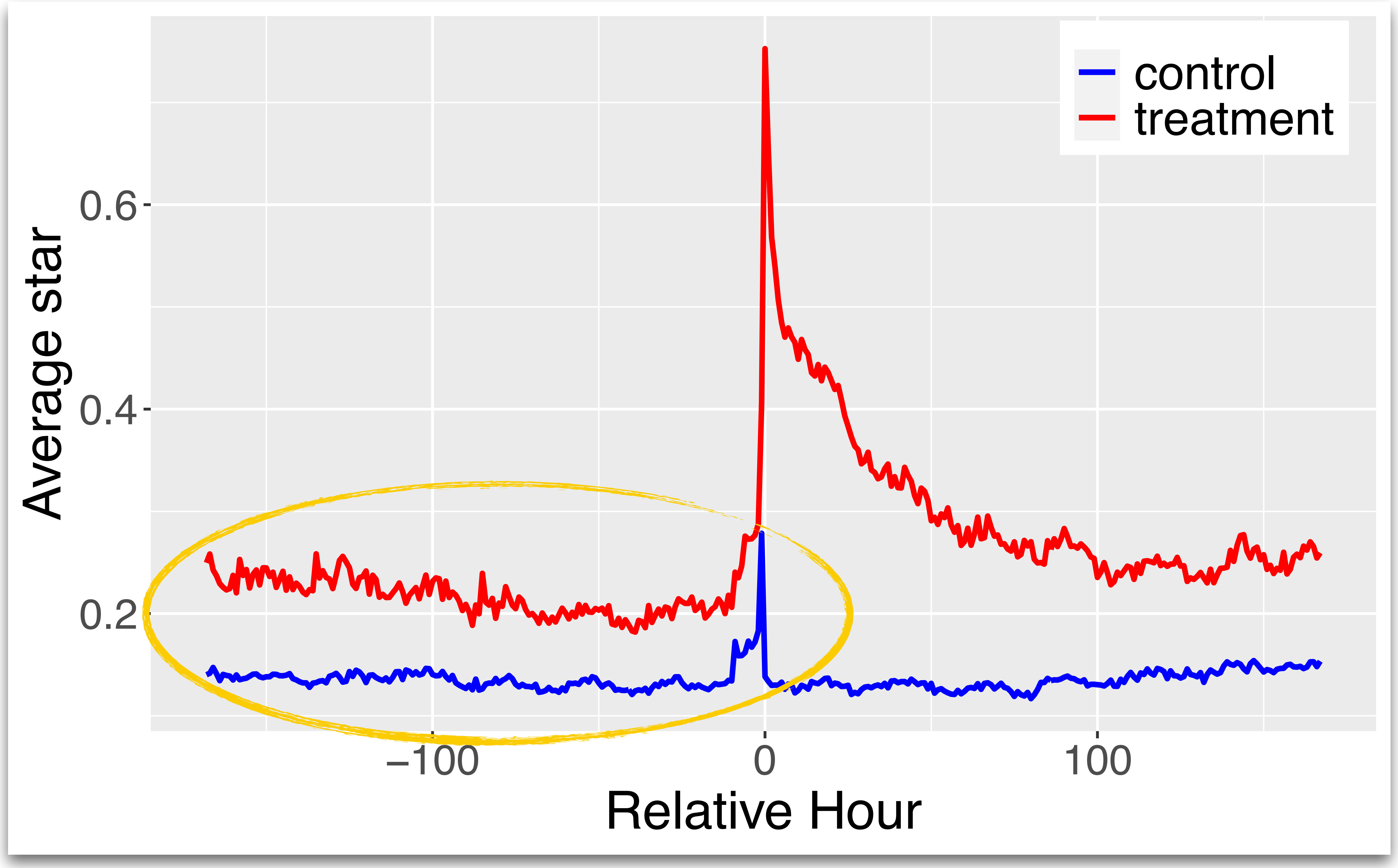
Carnegie Mellon University - S3D
Pittsburgh, PA
http://bvasiles.github.io
@b_vasilescu

Popular repositories

Repository	Visibility	Description	Language	Stars	Forks
empirical-methods	Public	Homepage for 17-803 "Empirical Methods" at Carnegie Mellon University	HTML	112	24
jsNaughty	Public	JS reverse minifier based on statistical machine translation	JavaScript	17	2
diversity	Public	A data set for social diversity studies of GitHub teams		13	4
ght_unmasking_aliases	Public		Python	11	9
bvasiles.github.io	Public	My website	HTML	3	1
SuffixTree	Public	Forked from JDonner/SuffixTree Continuation of Danny Yoo's SuffixTree and SubstringDict	C	1	

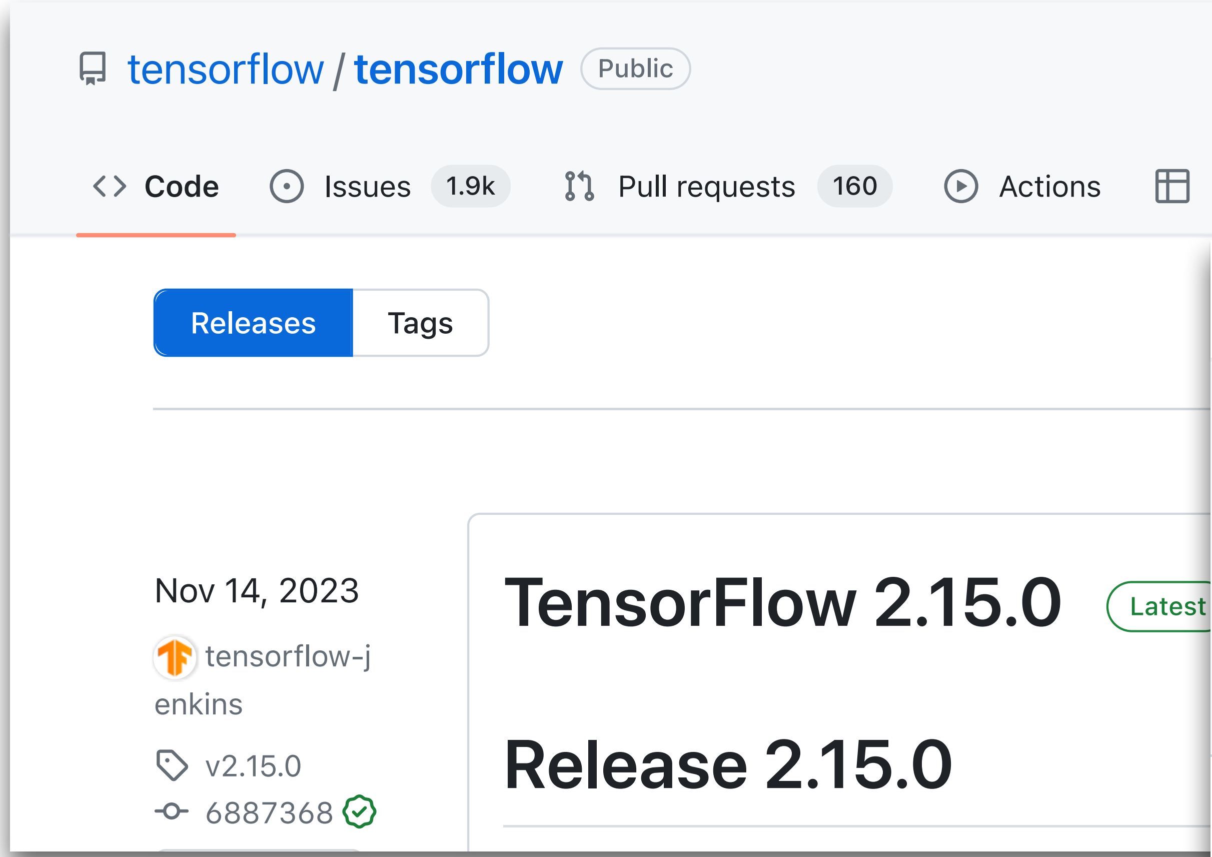
Challenge: Parallel trends assumption

Propensity score matching to ensure the control repositories, on average, have the same pre-treatment trend in outcome variables as the treatment group.



Challenge: Confounding events more likely to impact treatment group

Control for official releases, being featured on Trending, and overall Google chatter.



Nov 14, 2023

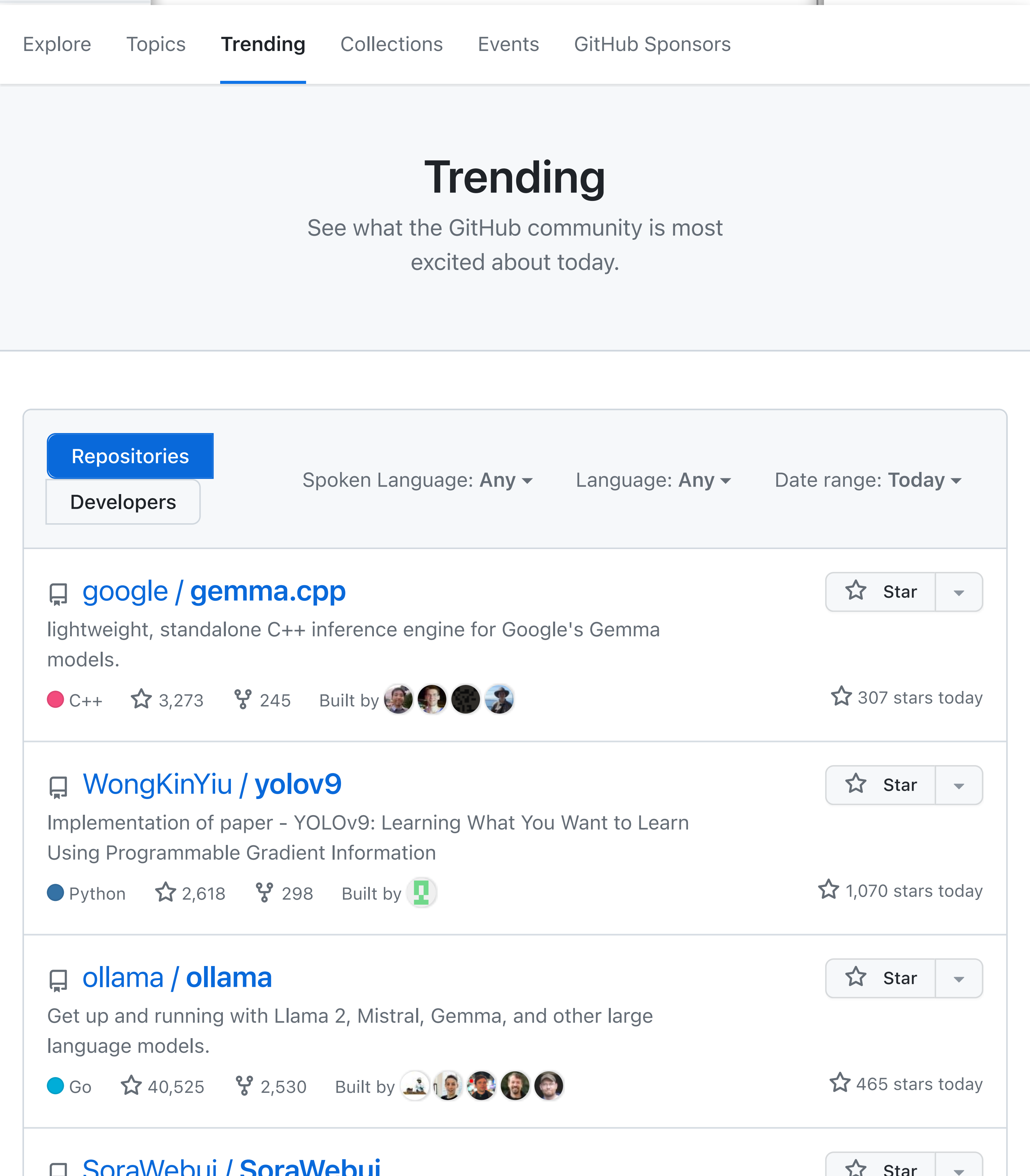
tensorflow-jenkins

v2.15.0

6887368

TensorFlow 2.15.0 Latest

Release 2.15.0



Trending

See what the GitHub community is most excited about today.

google / gemma.cpp

lightweight, standalone C++ inference engine for Google's Gemma models.

C++ 3,273 245 Built by — Star — 307 stars today

WongKinYiu / yolov9

Implementation of paper - YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information

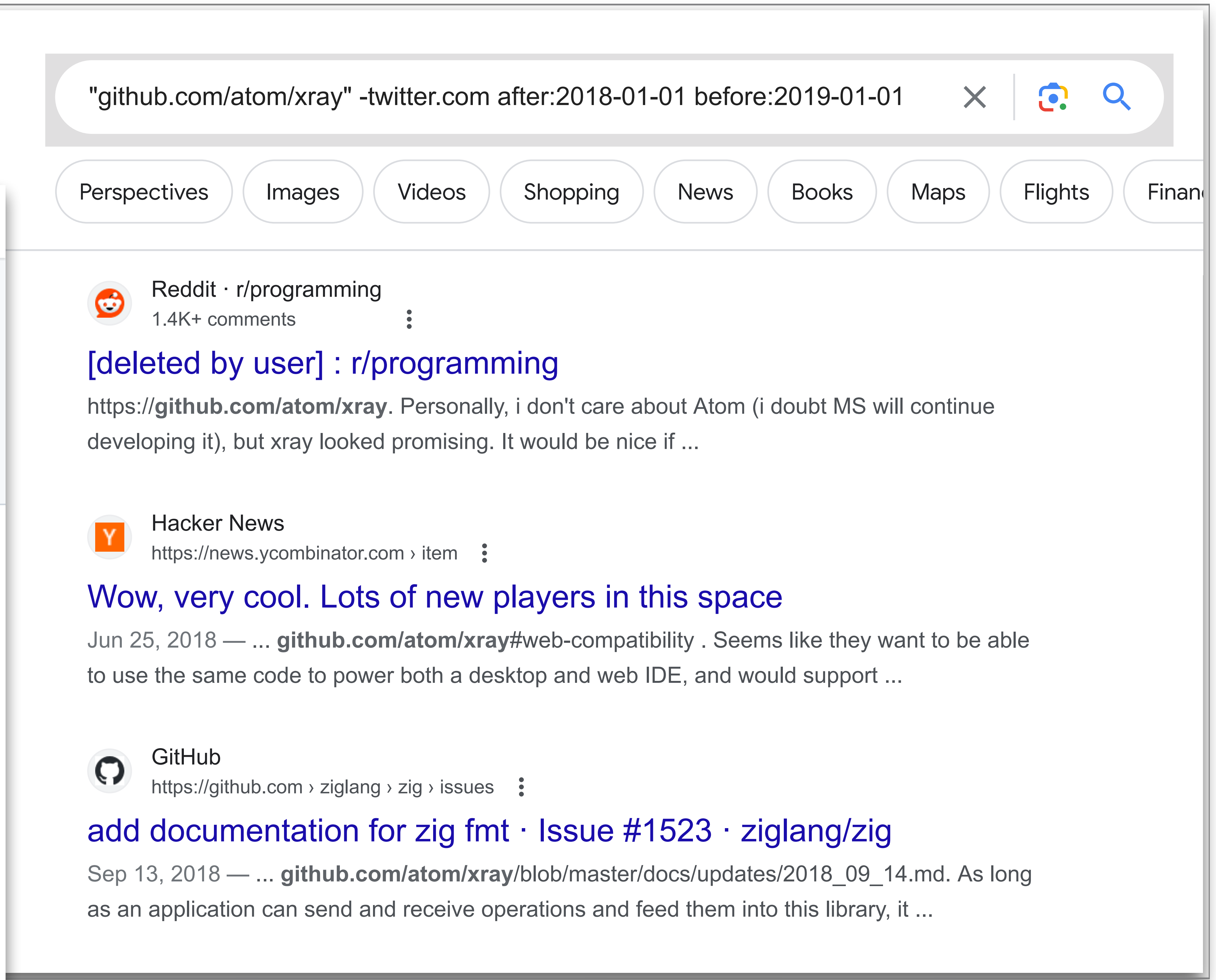
Python 2,618 298 Built by — Star — 1,070 stars today

ollama / ollama

Get up and running with Llama 2, Mistral, Gemma, and other large language models.

Go 40,525 2,530 Built by — Star — 465 stars today

SoraWebui / SoraWebui



"github.com/atom/xray" -twitter.com after:2018-01-01 before:2019-01-01

Reddit · r/programming
1.4K+ comments

[deleted by user] : r/programming

https://github.com/atom/xray. Personally, i don't care about Atom (i doubt MS will continue developing it), but xray looked promising. It would be nice if ...

Hacker News
https://news.ycombinator.com › item

Wow, very cool. Lots of new players in this space

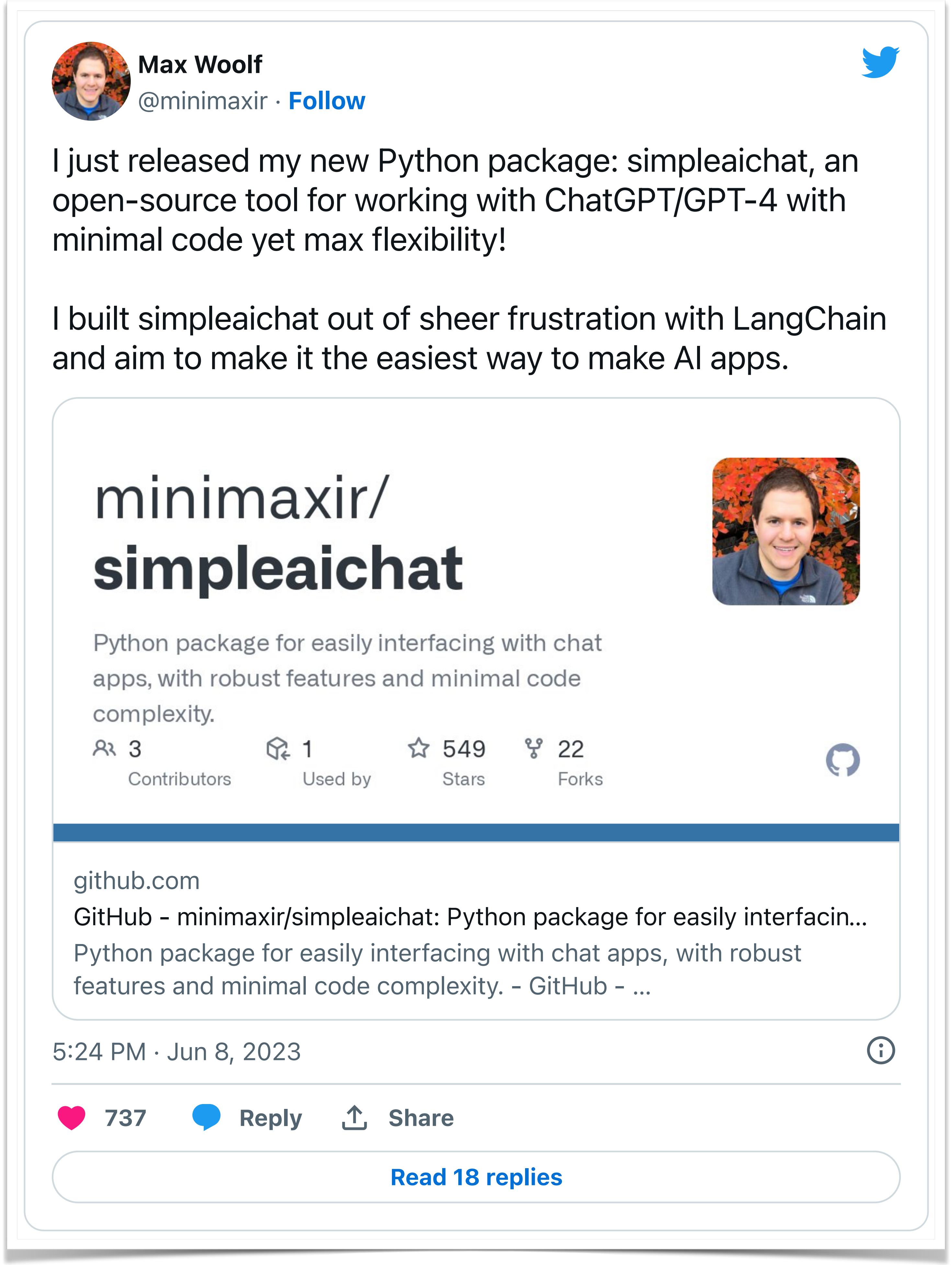
Jun 25, 2018 — ... github.com/atom/xray#web-compatibility . Seems like they want to be able to use the same code to power both a desktop and web IDE, and would support ...

GitHub
https://github.com › ziglang › zig › issues

add documentation for zig fmt · Issue #1523 · ziglang/zig

Sep 13, 2018 — ... github.com/atom/xray/blob/master/docs/updates/2018_09_14.md. As long as an application can send and receive operations and feed them into this library, it ...

Yes! Tweets cause GitHub stars and new contributors



Max Woolf
@minimaxir · Follow

I just released my new Python package: simpleaichat, an open-source tool for working with ChatGPT/GPT-4 with minimal code yet max flexibility!

I built simpleaichat out of sheer frustration with LangChain and aim to make it the easiest way to make AI apps.

minimaxir/
simpleaichat

Python package for easily interfacing with chat apps, with robust features and minimal code complexity.

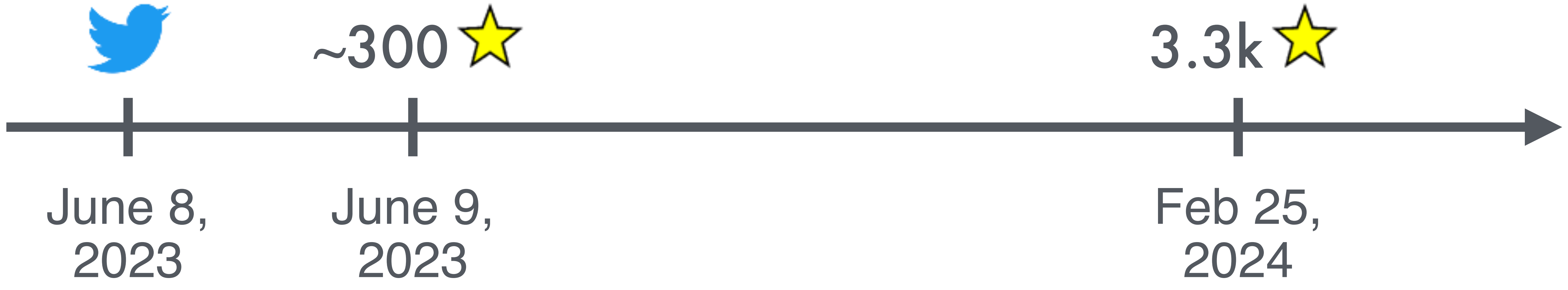
3 Contributors · 1 Used by · 549 Stars · 22 Forks

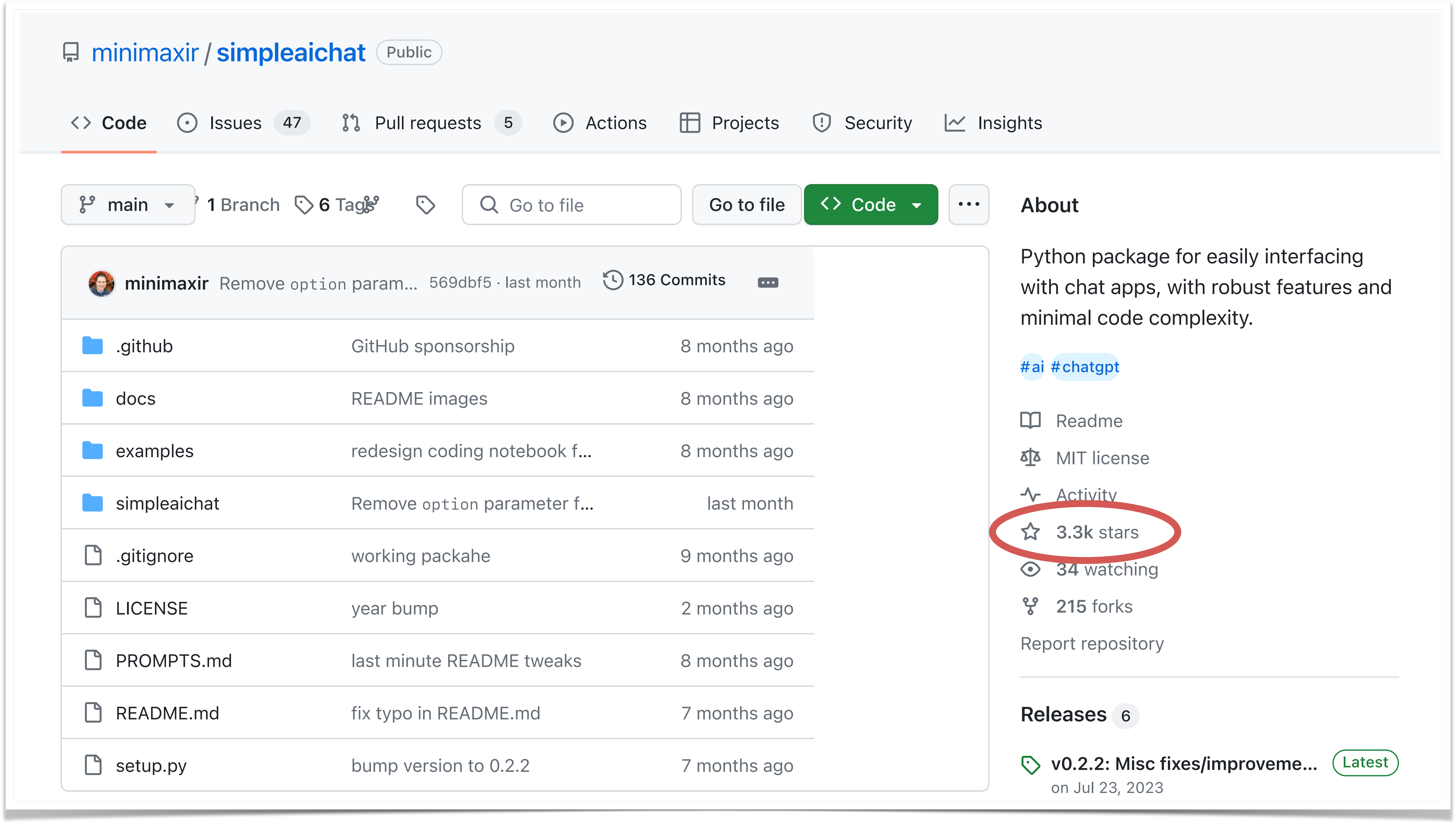
github.com
GitHub - minimaxir/simpleaichat: Python package for easily interfacin...
Python package for easily interfacing with chat apps, with robust features and minimal code complexity. - GitHub - ...

5:24 PM · Jun 8, 2023

737 · Reply · Share

Read 18 replies





minimaxir / simpleaichat Public

minimaxir Remove `option` param... 569dbf5 · last month · 136 Commits

.github	GitHub sponsorship	8 months ago
docs	README images	8 months ago
examples	redesign coding notebook f...	8 months ago
simpleaichat	Remove `option` parameter f...	last month
.gitignore	working packahe	9 months ago
LICENSE	year bump	2 months ago
PROMPTS.md	last minute README tweaks	8 months ago
README.md	fix typo in README.md	7 months ago
setup.py	bump version to 0.2.2	7 months ago

About

Python package for easily interfacing with chat apps, with robust features and minimal code complexity.

#ai #chatgpt

Readme
MIT license
Activity
3.3k stars
34 watching
215 forks
Report repository

Releases 6

v0.2.2: Misc fixes/improveme... Latest
on Jul 23, 2023

Dealing with abandoned upstream dependencies

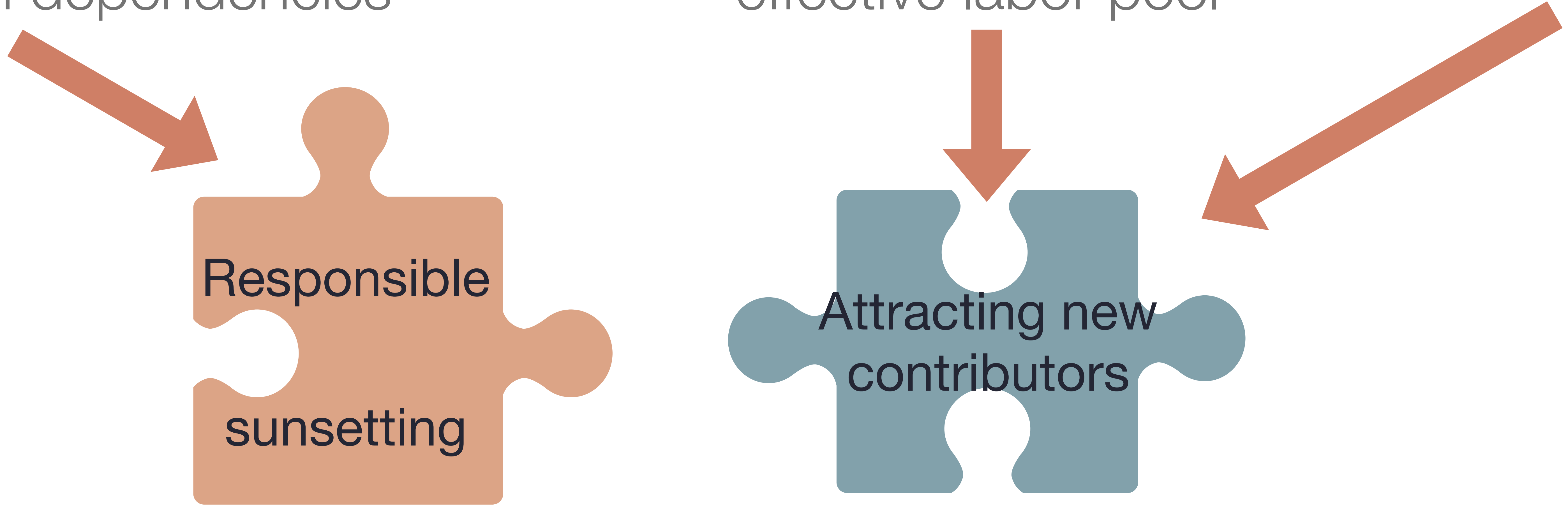
Estimating a project’s effective labor pool

Estimating causal effects of promotional activities

Dealing with abandoned upstream dependencies

Estimating a project’s effective labor pool

Estimating causal effects of promotional activities



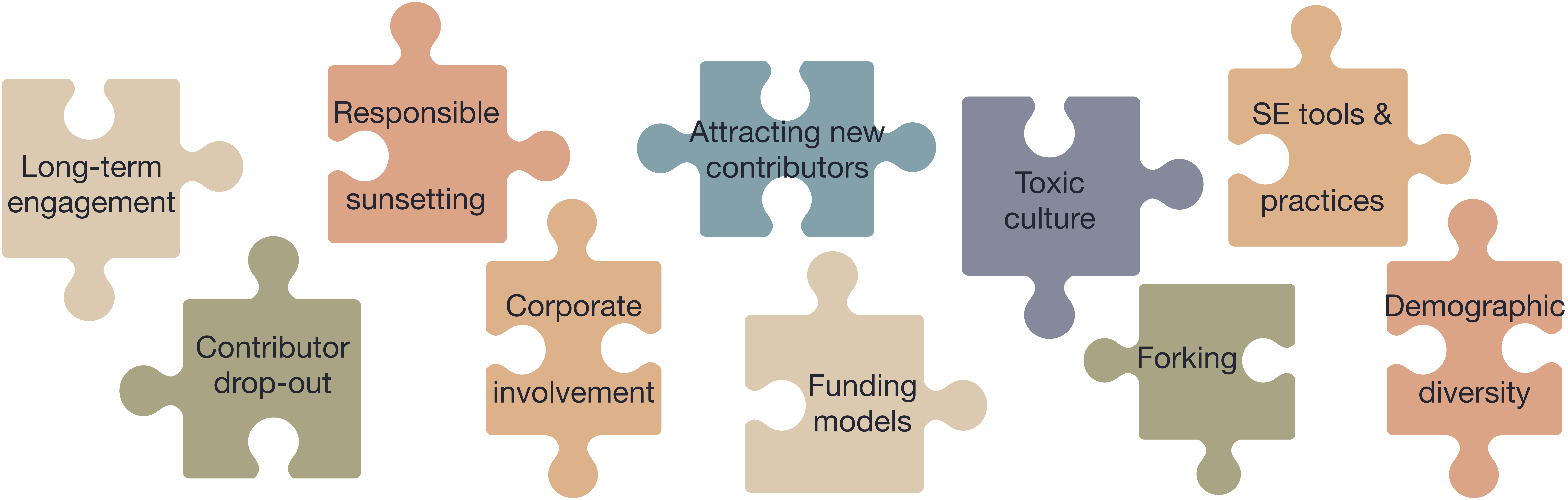
Responsible sunsetting

Attracting new contributors

Dealing with abandoned upstream dependencies

Estimating a project's effective labor pool

Estimating causal effects of promotional activities



Long-term engagement

Responsible sunsetting

Contributor drop-out

Corporate involvement

Attracting new contributors

Funding models

Toxic culture

SE tools & practices

Forking

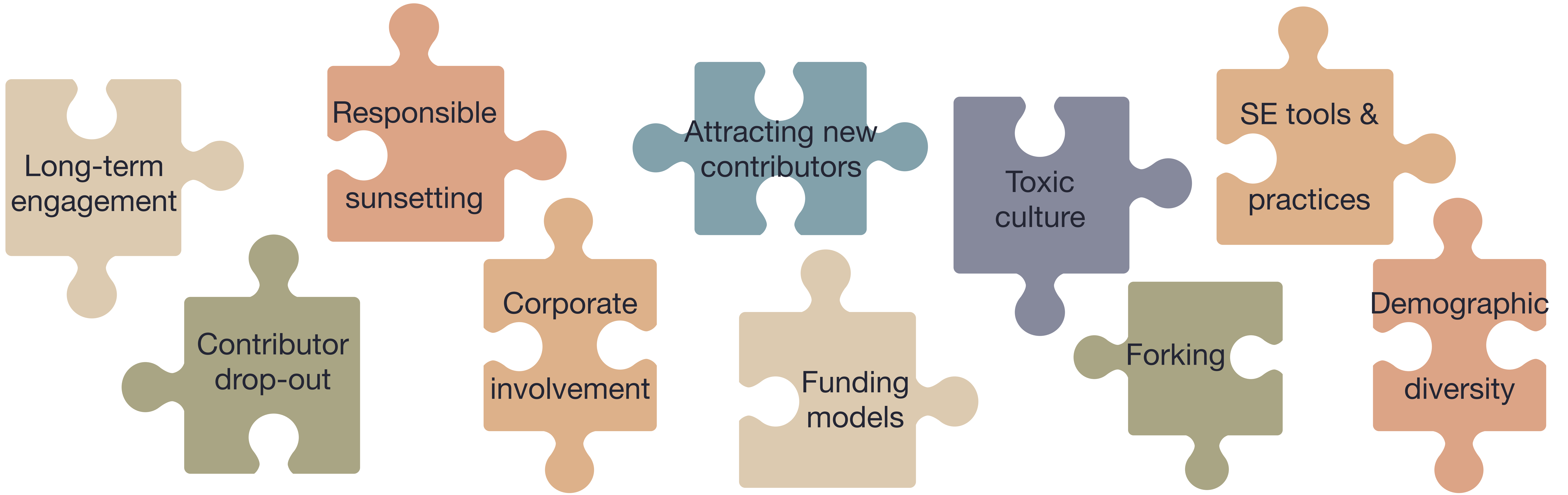
Demographic diversity

More open questions remaining than answers so far

- How does it all work?
  - How do the competing needs of different stakeholders get satisfied?
  - How does responsibility emerge?
- How healthy and sustainable is the ecosystem?

… especially with the attention it has been getting

- How to design effective interventions lacking centralized control?
- How do variations across contexts impact all of the above?



Long-term engagement

Responsible sunsetting

Attracting new contributors

Toxic culture

SE tools & practices

Contributor drop-out

Corporate involvement

Funding models

Forking

Demographic diversity

Acknowledgements

Courtney Miller

Anita Brown

Asher Trockman

Jim Herbsleb

Shurui Zhou

Hongbo Fang

Anita Sarma

Cassandra Overney

Audris Mockus

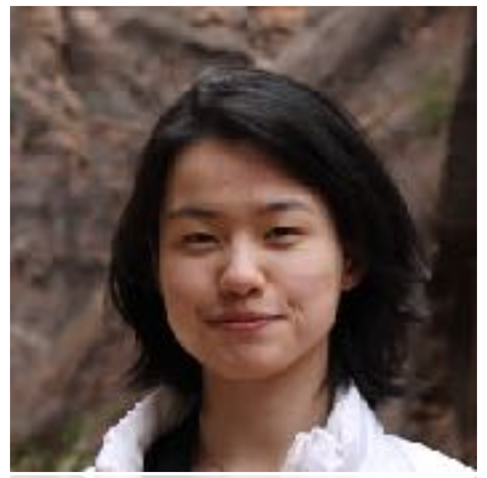
Alex Nolte

Sophie Qiu

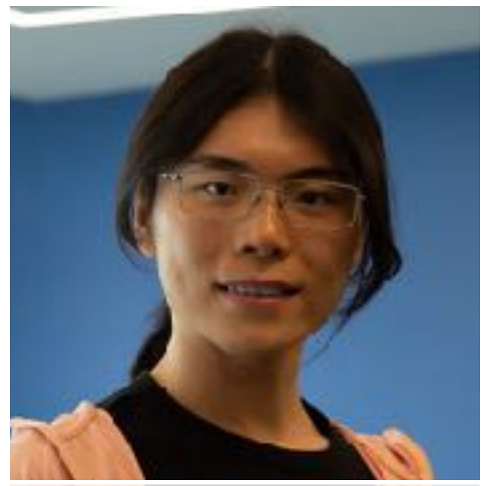
Alex Serebrenik

Marat Valiev

Laura Dabbish


Lily Li

Naveen Raman


Hao He

Christian Kästner

Hemank Lamba

Emerson Murphy-Hill

Google

Alfred P. Sloan FOUNDATION

National Science Foundation

FORDFOUNDATION

STRUDEL sustainability research on …

Project practices

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

Funding models

- ICSE 2020 (donations)

Sunsetting

- ESEC/FSE 2023
- ICSE 2025 (dealing with abandonment)

Attracting contributors

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

Transparency and signaling

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

Stress, burnout, disengagement

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

Diversity and inclusion

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

Novelty and innovation

- ICSE 2024 (atypical combinations)

Network effects

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