OF TEAMS AND AUTOMATION

LESSONS IN SOCIAL CODING FROM GITHUB

Bogdan Vasilescu

OF TEAMS AND AUTOMATION

LESSONS IN SOCIAL CODING FROM GITHUB

Bogdan Vasilescu

with thanks to

Baishakhi Ray

Alexander Serebrenik

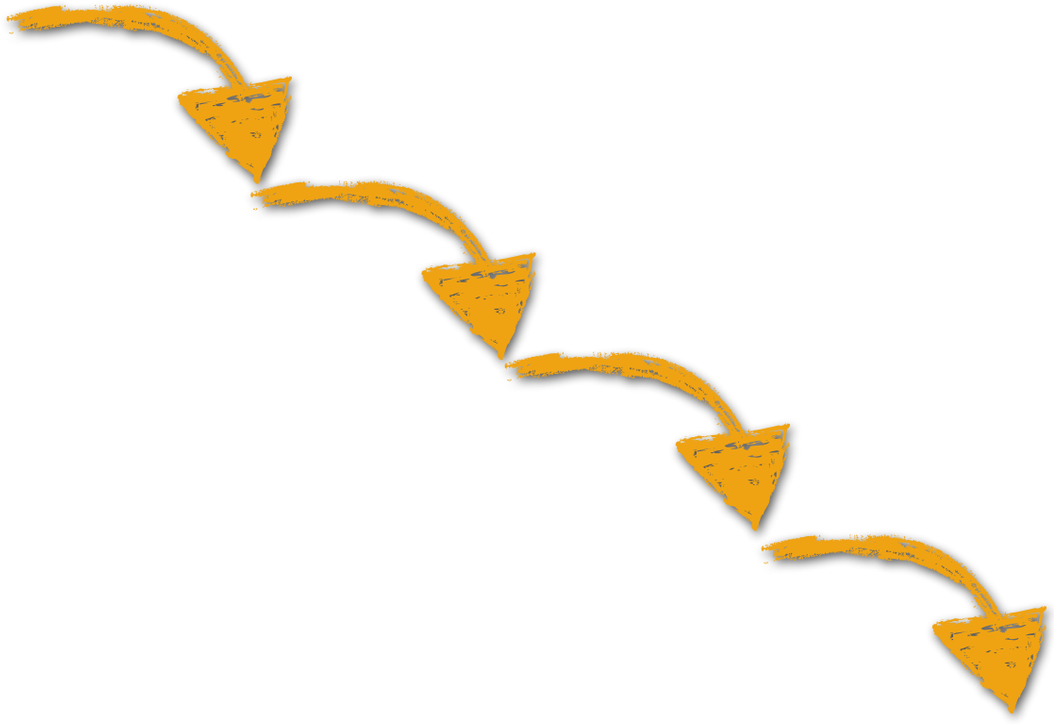
Vladimir Filkov

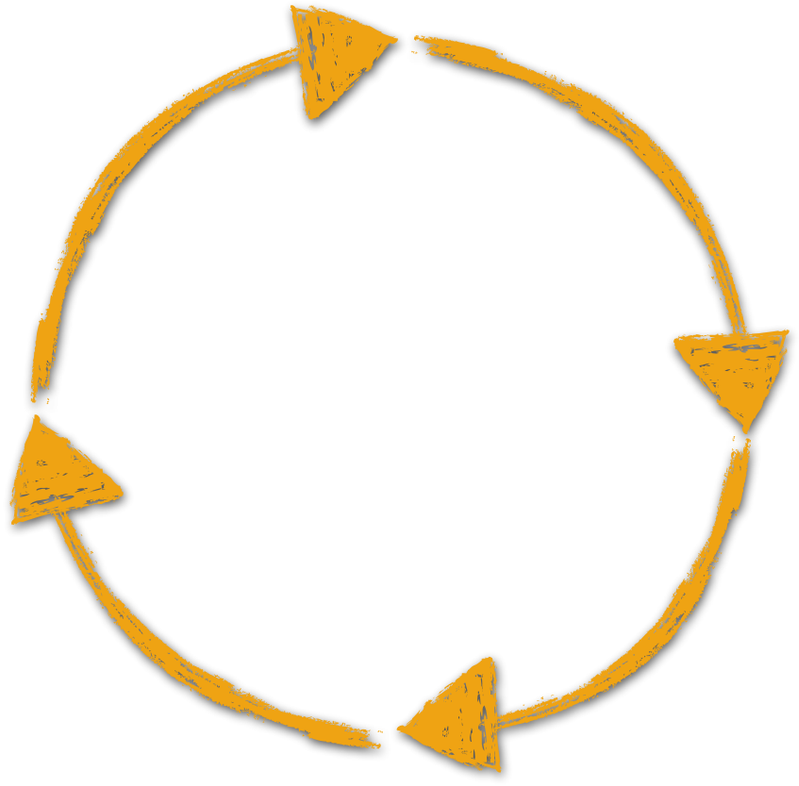
Prem Devanbu

Daryl Posnett

Yue Yu

Software development 2.0



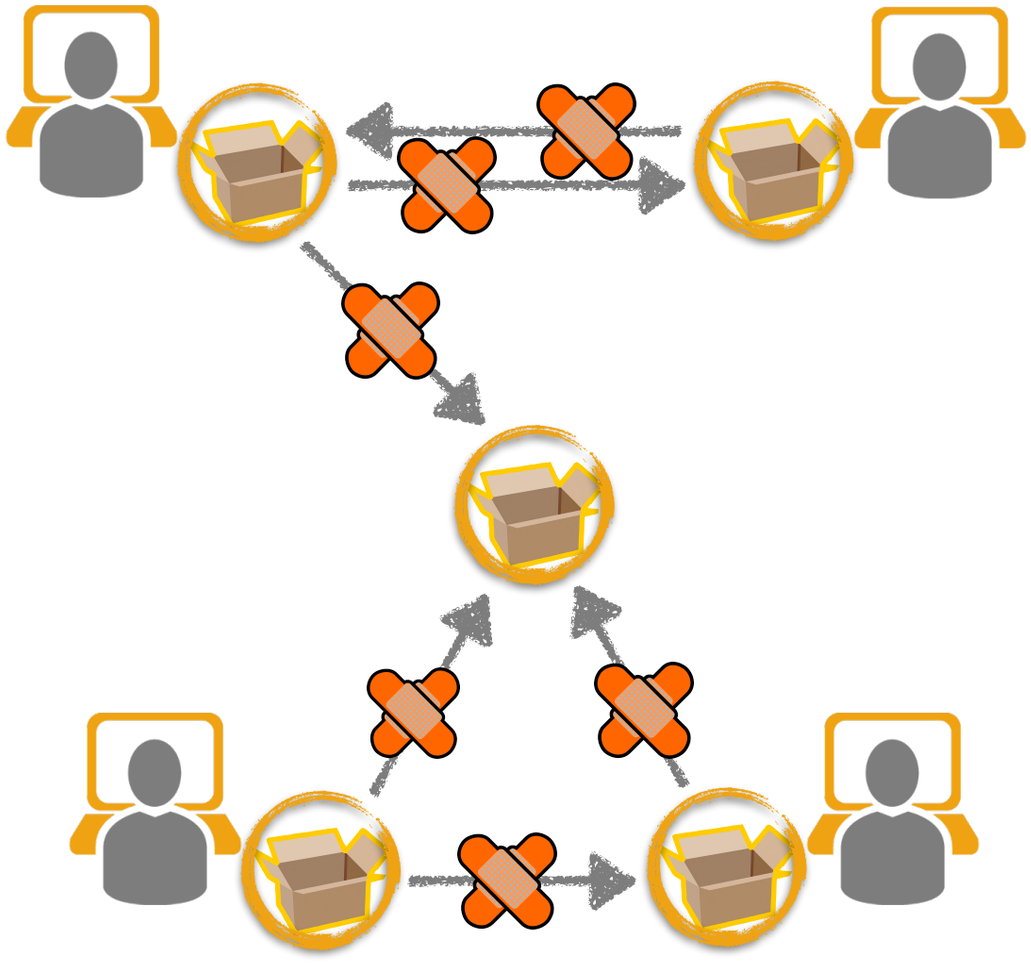


Waterfall vs. Agile

Software development 2.0

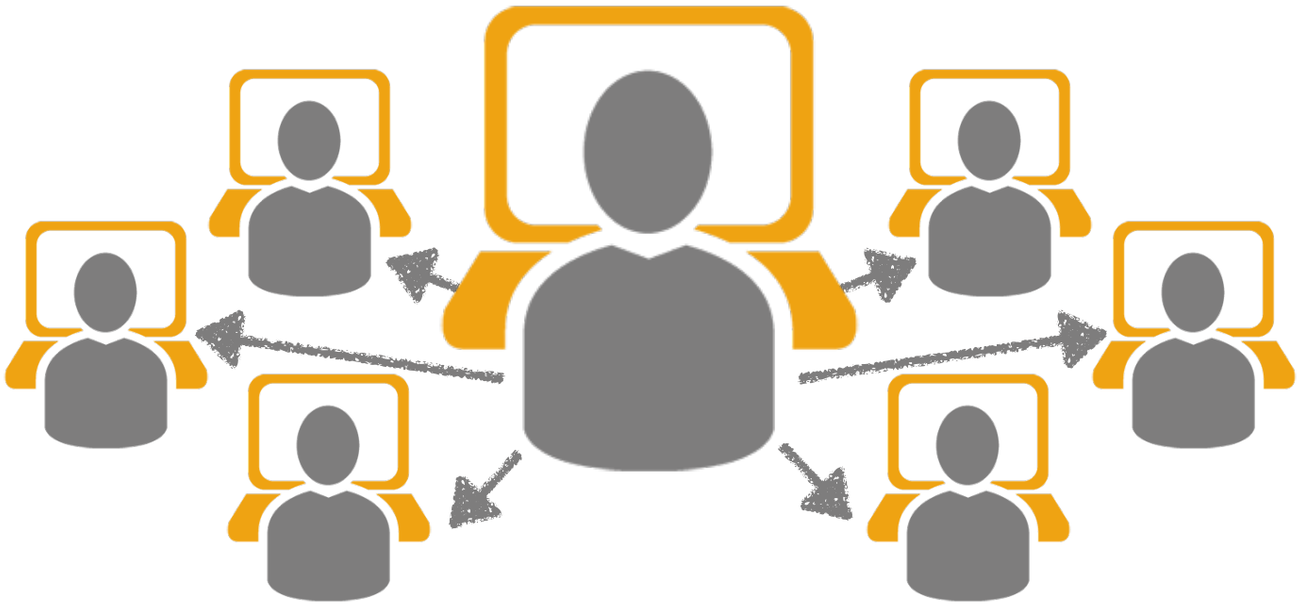
svn

vs.



git

The “social programmer”



stackoverflow

LinkedIn

github
SOCIAL CODING

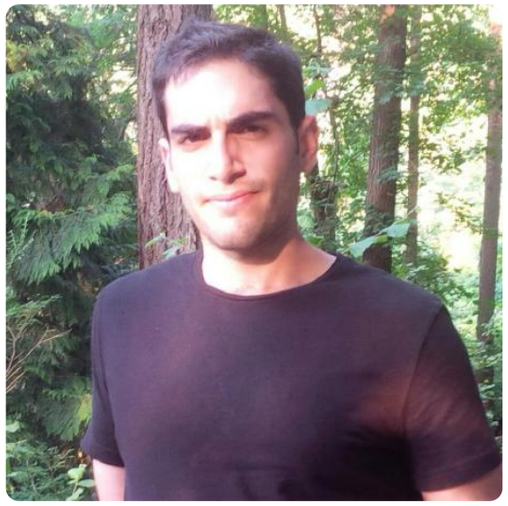
coderwall

Masterbranch

twitter

- Programming in a socially networked world: the evolution of the social programmer
  C Treude, F Figueira Filho, B Cleary, MA Storey. *The Future of Collaborative Software Development @ CSCW 2012*
- Social networking meets software development: Perspectives from GitHub, MSDN, Stack Exchange, and TopCoder
  A Begel, J Bosch, MA Storey. *IEEE Software 2013*

… has a public profile



Daniel Rosenwasser

DanielRosenwasser

Microsoft
Seattle, WA
https://bitbucket.org/DanRose…
Joined on Aug 10, 2011

38 Followers **22** Starred **14** Following

Organizations

Contributions | Repositories | Public activity

Follow

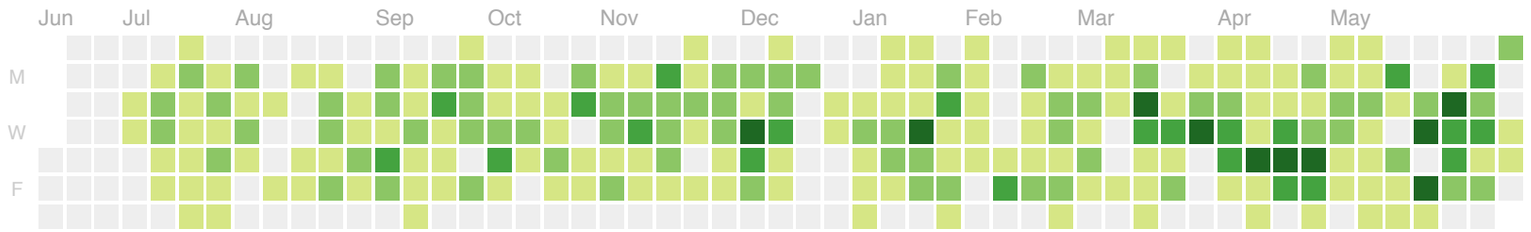
Popular repositories

- **mlton** — The MLton repository — 1 ★
- **TypeScript** — TypeScript is a superset of JavaScript that co… — 1 ★
- **bond** — Bond is a cross-platform framework for workin… — 0 ★
- **compat-table** — ECMAScript 5/6 compatibility tables — 0 ★
- **coreclr** — This repo contains the .NET Core runtime, call… — 0 ★

Repositories contributed to

- Microsoft/**TypeScript** — TypeScript is a superset of JavaScript that co… — 5,630 ★
- Microsoft/**TypeScript-wiki** — A repository to make changes to the TypeScri… — 23 ★
- Microsoft/**TypeScript-Handbook** — The TypeScript Handbook is a comprehensive… — 38 ★
- kangax/**compat-table** — ECMAScript 5/6 compatibility tables — 560 ★
- Microsoft/**TypeScriptSamples** — Samples for TypeScript — 320 ★

Public contributions



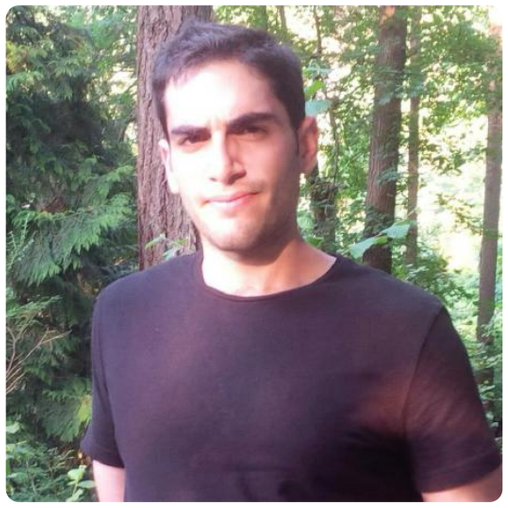
Summary of Pull Requests, issues opened, and commits. Learn more.

Less More

Contributions in the last year	Longest streak	Current streak
1,365 total	12 days	2 days
Jun 19, 2014 – Jun 19, 2015	April 27 – May 8	June 17 – June 18

- Social coding in GitHub: transparency and collaboration in an open software repository
  L Dabbish, C Stuart, J Tsay, J Herbsleb. *CSCW 2012*

… shares code



Daniel Rosenwasser
DanielRosenwasser

Microsoft
Seattle, WA
https://bitbucket.org/DanRose…
Joined on Aug 10, 2011

38 Followers **22** Starred **14** Following

Organizations

Contributions | Repositories | Public activity — Follow

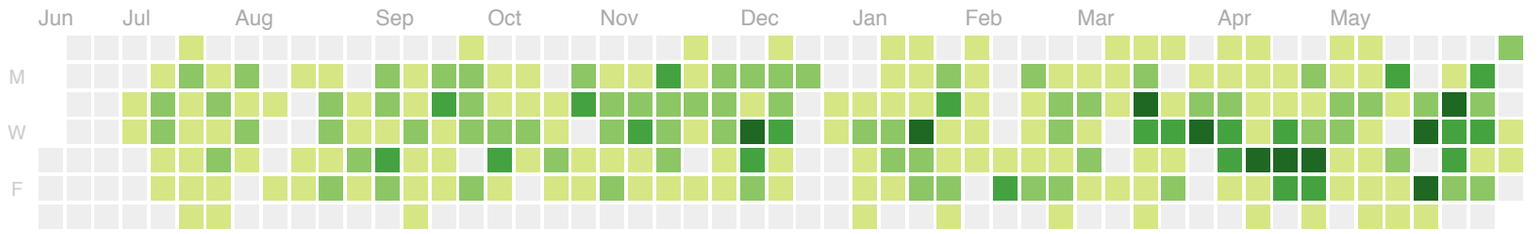
Popular repositories

- **mlton** — The MLton repository — 1 ★
- **TypeScript** — TypeScript is a superset of JavaScript that co… — 1 ★
- **bond** — Bond is a cross-platform framework for workin… — 0 ★
- **compat-table** — ECMAScript 5/6 compatibility tables — 0 ★
- **coreclr** — This repo contains the .NET Core runtime, call… — 0 ★

Repositories contributed to

- Microsoft/**TypeScript** — TypeScript is a superset of JavaScript that co… — 5,630 ★
- Microsoft/**TypeScript-wiki** — A repository to make changes to the TypeScri… — 23 ★
- Microsoft/**TypeScript-Handbook** — The TypeScript Handbook is a comprehensive… — 38 ★
- kangax/**compat-table** — ECMAScript 5/6 compatibility tables — 560 ★
- Microsoft/**TypeScriptSamples** — Samples for TypeScript — 320 ★

Public contributions

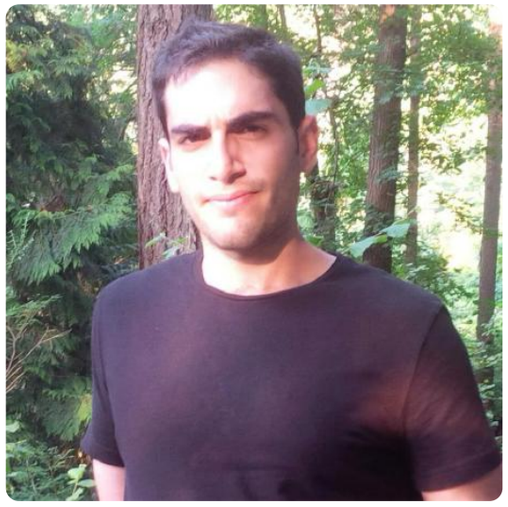


Summary of Pull Requests, issues opened, and commits. Learn more. Less More

Contributions in the last year	Longest streak	Current streak
1,365 total	12 days	2 days
Jun 19, 2014 – Jun 19, 2015	April 27 – May 8	June 17 – June 18

- Social coding in GitHub: transparency and collaboration in an open software repository
  L Dabbish, C Stuart, J Tsay, J Herbsleb. *CSCW 2012*

… is social



Daniel Rosenwasser

DanielRosenwasser

Microsoft
Seattle, WA
https://bitbucket.org/DanRose…
Joined on Aug 10, 2011

38 Followers **22** Starred **14** Following

Organizations

Contributions | Repositories | Public activity

Follow

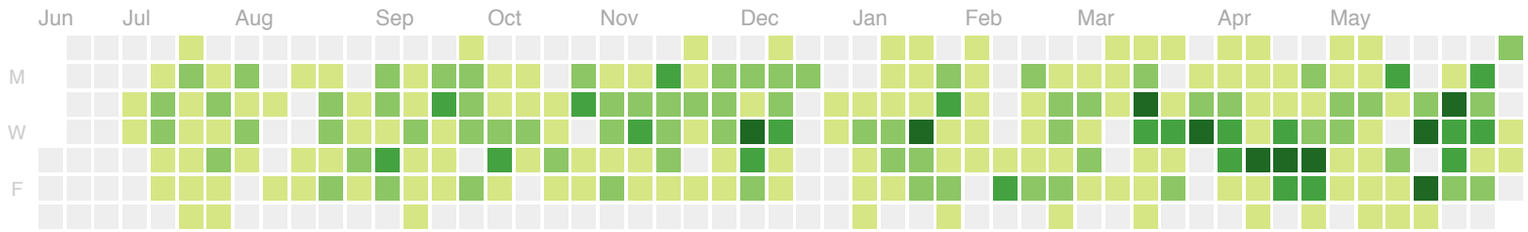
Popular repositories

- **mlton** — The MLton repository — 1 ★
- **TypeScript** — TypeScript is a superset of JavaScript that co… — 1 ★
- **bond** — Bond is a cross-platform framework for workin… — 0 ★
- **compat-table** — ECMAScript 5/6 compatibility tables — 0 ★
- **coreclr** — This repo contains the .NET Core runtime, call… — 0 ★

Repositories contributed to

- Microsoft/**TypeScript** — TypeScript is a superset of JavaScript that co… — 5,630 ★
- Microsoft/**TypeScript-wiki** — A repository to make changes to the TypeScri… — 23 ★
- Microsoft/**TypeScript-Handbook** — The TypeScript Handbook is a comprehensive… — 38 ★
- kangax/**compat-table** — ECMAScript 5/6 compatibility tables — 560 ★
- Microsoft/**TypeScriptSamples** — Samples for TypeScript — 320 ★

Public contributions



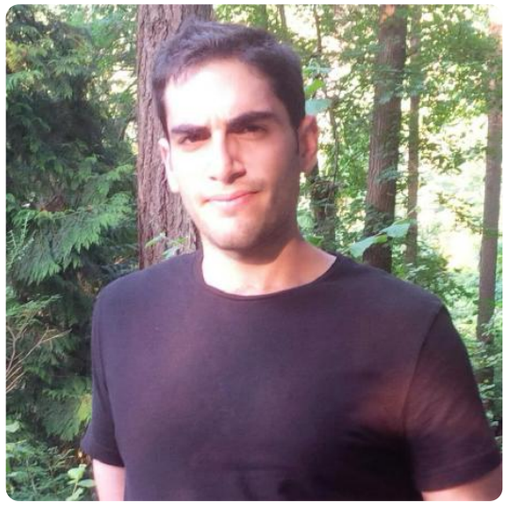
Summary of Pull Requests, issues opened, and commits. Learn more.

Less More

Contributions in the last year	Longest streak	Current streak
1,365 total	12 days	2 days
Jun 19, 2014 – Jun 19, 2015	April 27 – May 8	June 17 – June 18

- Social coding in GitHub: transparency and collaboration in an open software repository
  L Dabbish, C Stuart, J Tsay, J Herbsleb. *CSCW 2012*

… signals



Daniel Rosenwasser

DanielRosenwasser

Microsoft
Seattle, WA
https://bitbucket.org/DanRose…
Joined on Aug 10, 2011

38 Followers **22** Starred **14** Following

Organizations

Contributions | Repositories | Public activity

Follow

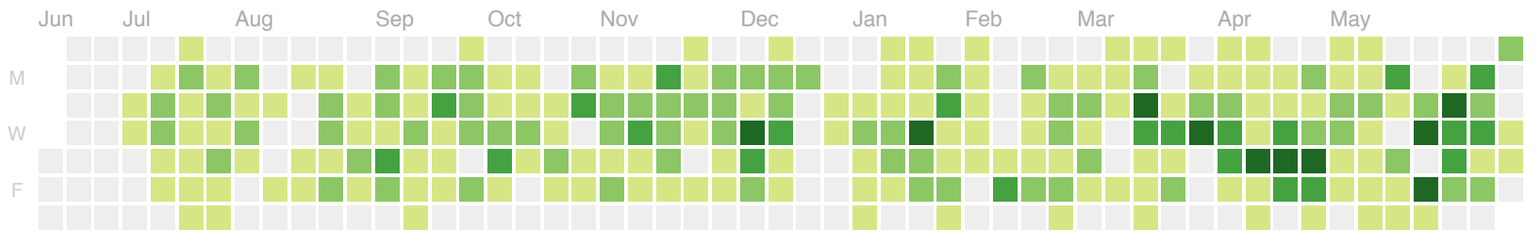
Popular repositories

- **mlton** — The MLton repository — 1 ★
- **TypeScript** — TypeScript is a superset of JavaScript that co… — 1 ★
- **bond** — Bond is a cross-platform framework for workin… — 0 ★
- **compat-table** — ECMAScript 5/6 compatibility tables — 0 ★
- **coreclr** — This repo contains the .NET Core runtime, call… — 0 ★

Repositories contributed to

- Microsoft/**TypeScript** — TypeScript is a superset of JavaScript that co… — 5,630 ★
- Microsoft/**TypeScript-wiki** — A repository to make changes to the TypeScri… — 23 ★
- Microsoft/**TypeScript-Handbook** — The TypeScript Handbook is a comprehensive… — 38 ★
- kangax/**compat-table** — ECMAScript 5/6 compatibility tables — 560 ★
- Microsoft/**TypeScriptSamples** — Samples for TypeScript — 320 ★

Public contributions



Summary of Pull Requests, issues opened, and commits. Learn more.

Less More

Contributions in the last year	Longest streak	Current streak
1,365 total	12 days	2 days
Jun 19, 2014 – Jun 19, 2015	April 27 – May 8	June 17 – June 18

- Social coding in GitHub: transparency and collaboration in an open software repository
  L Dabbish, C Stuart, J Tsay, J Herbsleb. *CSCW 2012*

… shares knowledge

784,785 REPUTATION

360 5367 6604

Jon Skeet top 0.01% overall

Author of C# in Depth.
Currently a software engineer at Google, London.
Usually a Microsoft MVP (C#, 2003-2010, 2011-)

Sites:

- C# in Depth
- Coding blog
- C# articles
- Twitter updates (@jonskeet)
- Google+ profile

Email: skeet@pobox.com (but please read my blog post on Stack Overflow-related emails first)

31,797 answers **34** questions **~118.2m** people reached

Reading, United Kingdom
jonskeet
jskeet
csharpindepth.com
Member for 6 years, 8 months
978,919 profile views
Last seen 7 mins ago

Communities (15)

Stack Overflow	784.8k
Meta Stack Exchange	70.5k
Super User	4.1k
Programmers	3k
Server Fault	2.8k

View network profile →

Top Tags (4,445)

c#	SCORE 145,166	POSTS 17,262	POSTS % 54
java	SCORE 85,729	POSTS 9,627	
.net	SCORE 50,165	POSTS 5,145	
linq	SCORE 20,239	POSTS 2,717	
string	SCORE 11,990	POSTS 916	
generics	SCORE 11,511	POSTS 1,161	

View all tags →

Top Meta Posts

1 9

320 Stack Overflow question checklist

Top Posts (31,831)

All | Questions | Answers | Votes | Newest

6272 Why is subtracting these two times (in 1927) giving a strange result? jul 27 '11

… signals

Stack Overflow

Q&A for professional and enthusiast programmers

(4,368,896 total users)

All Time

Type to find users:

Jon Skeet

360 5367 6604

member for: 6 years, 8 months

#1 all time rank

784,585 all time reputation

Darin Dimitrov

98 2027 2062

member for: 6 years, 8 months

#2 all time rank

595,006 all time reputation

BalusC

142 1872 2174

member for: 5 years, 10 months

#3 all time rank

573,150 all time reputation

Hans Passant

61 656 1236

member for: 6 years, 9 months

#4 all time rank

553,678 all time reputation

… builds reputation

784,785 REPUTATION

360 5367 6604

Jon Skeet top 0.01% overall

31,797 answers 34 questions ~118.2m people reached

Author of C# in Depth

- C# in Depth
- Coding blog
- C# article
- Twitter u
- Google+

Email: skeet@

Overflow-relat

jskeet

Communities (15)

- Stack Overflow 784.8k
- Meta Stack Exchange 70.5k
- Super User 4.1k
- Programmers 3k
- Server Fault 2.8k

View network profile →

Top Meta Posts 1 9

320 Stack Overflow question checklist

linq

Top Posts

6272 Why is subtracting these two times (in 1927) giving a strange result?

Jul 27 '11

54

,145

- "Jon Skeet **can divide by zero**."
- "When Jon Skeet's code fails to compile the **compiler apologises**."
- "Jon Skeet does not use revision control software. **None of his code has ever needed revision**."
- "Jon Skeet has already written a book about C# 6.0. It's currently sealed up. In three years, Anders Hejlsberg is going to open the book to **see if the language design team got it right**."

http://meta.stackexchange.com/questions/9134/jon-skeet-facts

Impact?

784,785 REPUTATION

360 5367 6604

Communities (15)

- Stack Overflow 784.8k
- Meta Stack Exchange 70.5k
- Super User 4.1k
- Programmers 3k
- Server Fault 2.8k

View network profile →

Top Meta Posts 1 9

320 Stack Overflow question checklist

Jon Skeet top 0.01% overall

Author of C# in Depth.
Currently a software engineer at Google, London.
Usually a Microsoft MVP (C#, 2003-2010, 2011-)

Sites:

- C# in Depth
- Coding blog
- C# articles
- Twitter updates (@jonskeet)
- Google+ profile

Email: skeet@pobox.com (but please read my blog post on S
Overflow-related emails first)

31,797 answers | 34 questions | ~118.2m people reached

Reading, United Kingdom

Last seen 7 mins ago

~118.2m
people reached

Top Tags (4,445)

c# java linq

POSTS 2,717 POSTS 916 POSTS 1,161

View all tags →

2014 IDC world estimates:

- 29m ICT-skilled workers
- 18.5m software developers (professional + hobbyist)

Top Posts (31,831) All | Questions | Answers | Votes | Newest

6272 Why is subtracting these two times (in 1927) giving a strange result? jul 27 '11

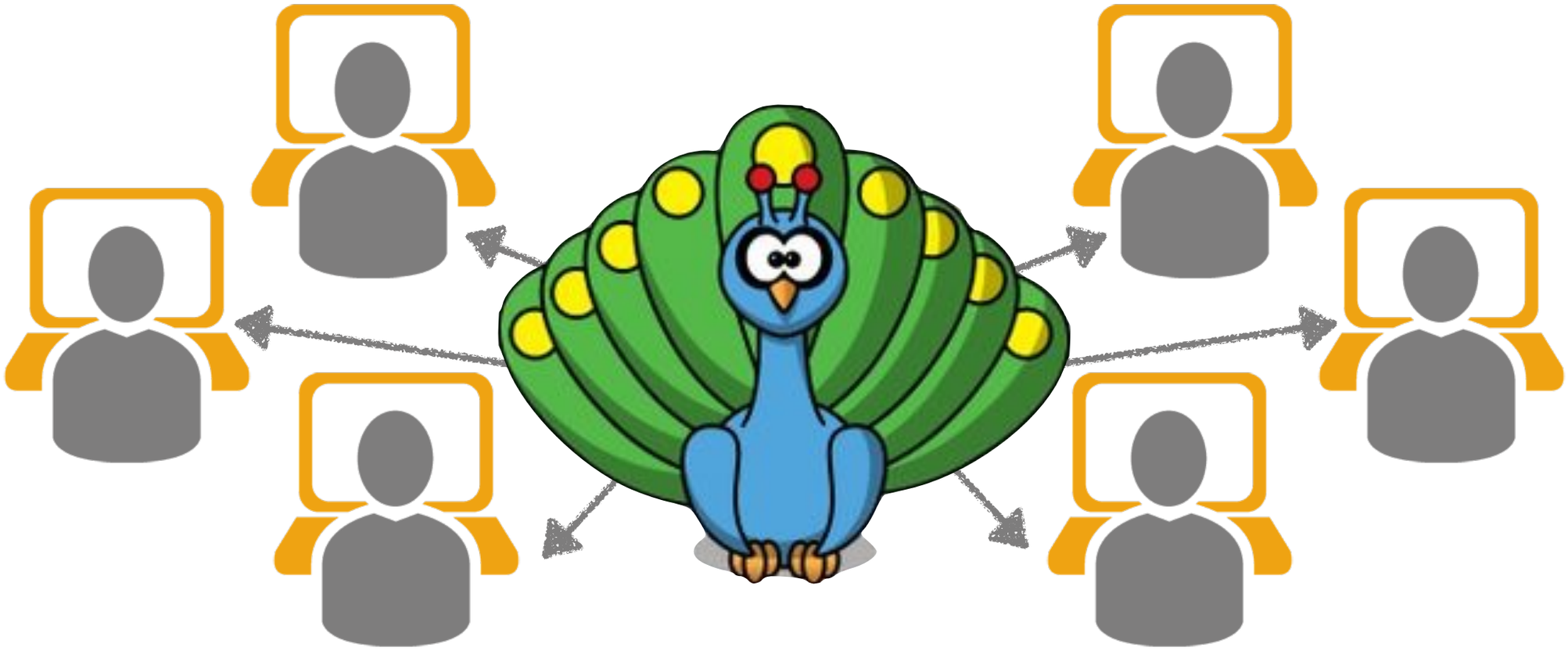
Impact?

Masterbranch

“brings all your projects together in one neat profile and let you **show off** your coding skills to developers and employers”

The “social programmer”

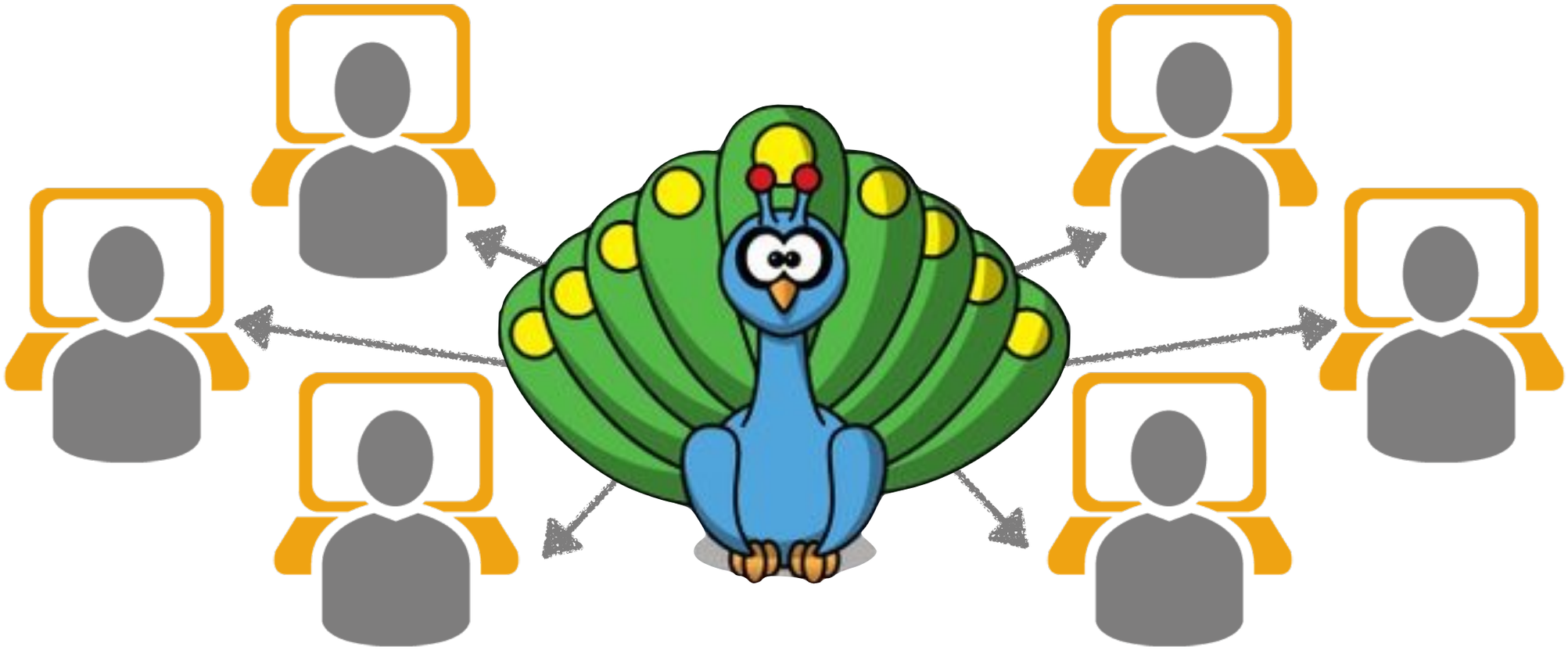
… signals



- Assessing technical candidates on the social web
  A Capiluppi, A Serebrenik, L Singer. *IEEE Software 2013*
- Mutual assessment in the social programmer ecosystem: an empirical investigation of developer profile aggregators
  L Singer, F Figueira Filho, B Cleary, C Treude, MA Storey, K Schneider. *CSCW 2013*
- Impression formation in online peer production: activity traces and personal profiles in GitHub
  J Marlow, L Dabbish, J Herbsleb. *CSCW 2013*
- Activity traces and signals in software developer recruitment and hiring
  J Marlow, L Dabbish. *CSCW 2013*

The “social programmer”

… signals



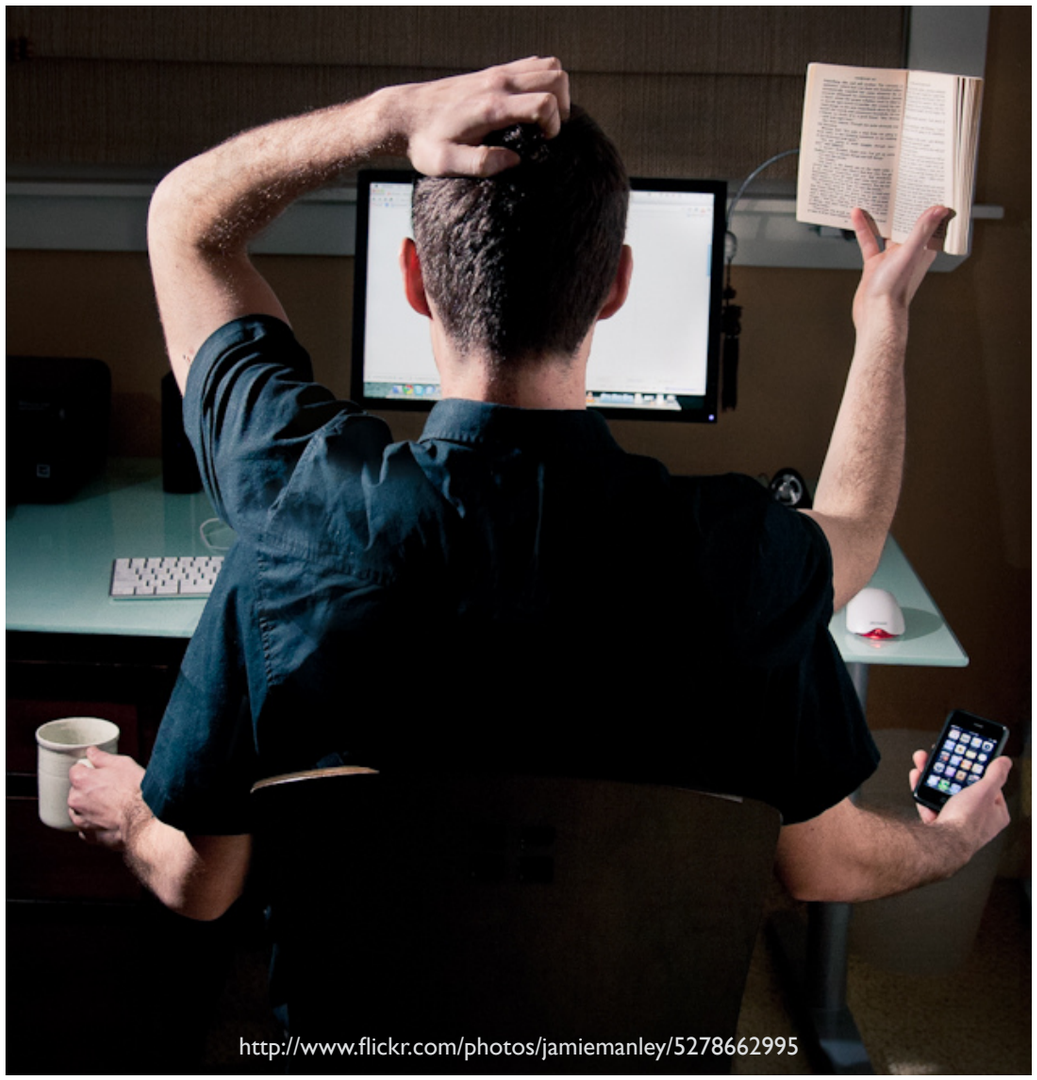
GitHub

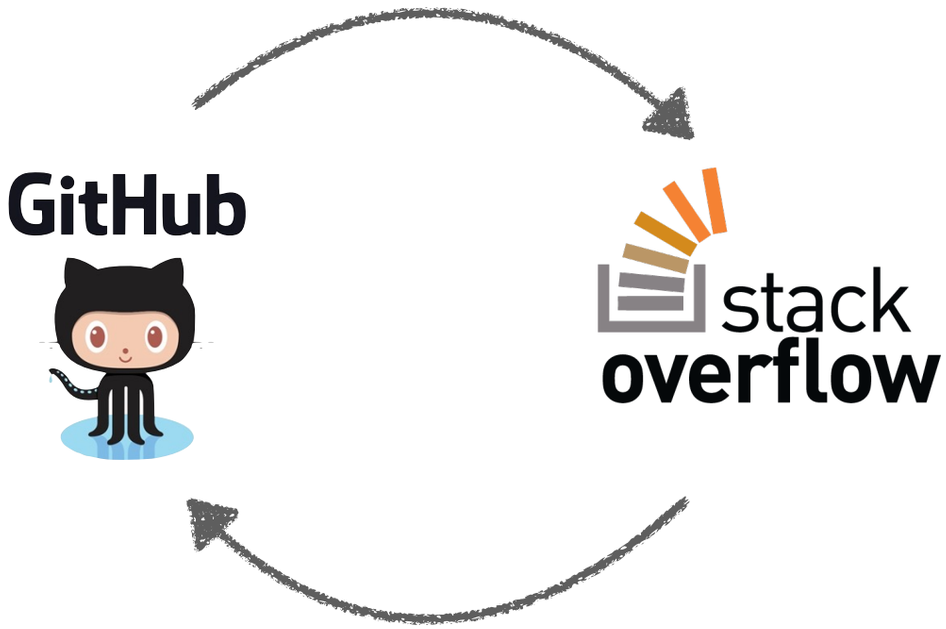
stack overflow

MAILING LIST

Stack Exchange

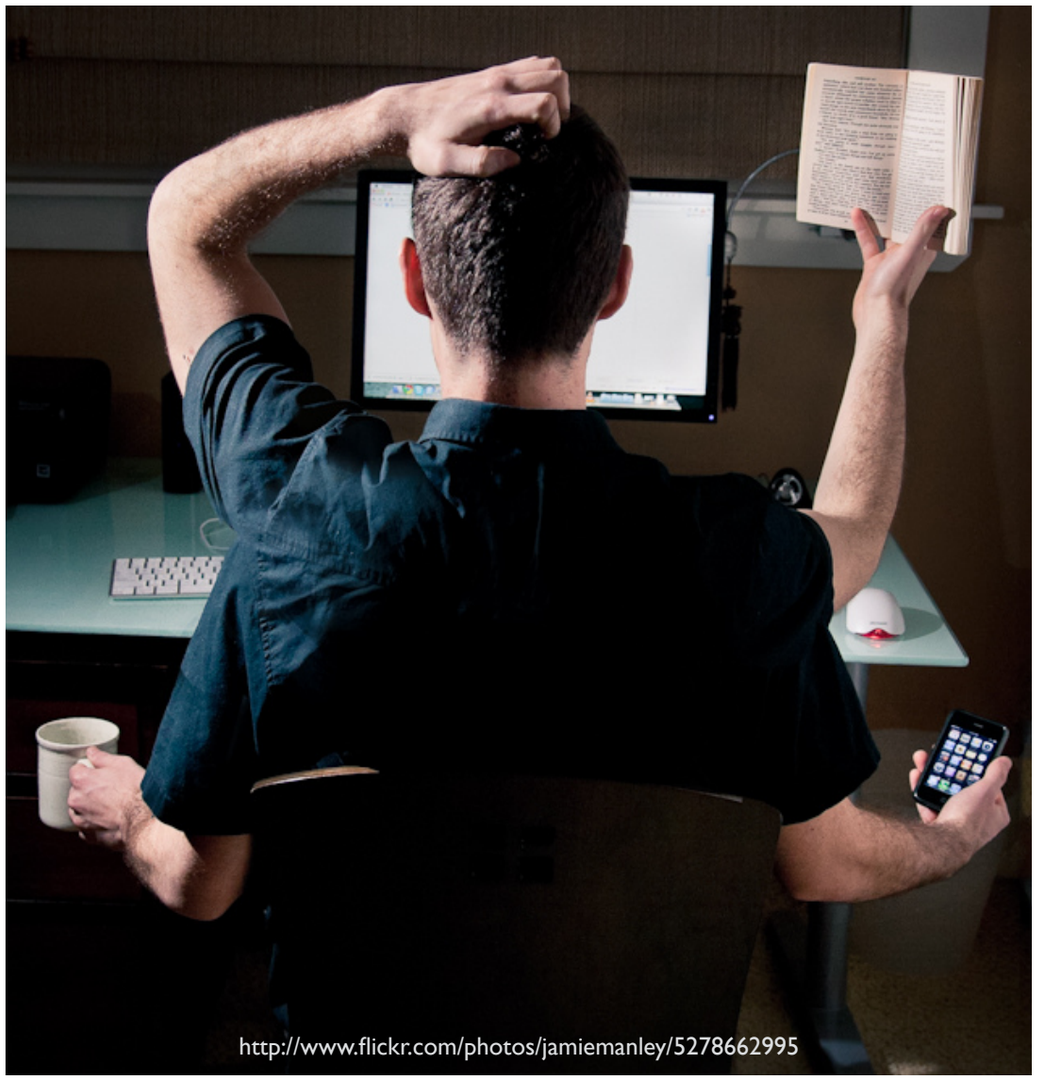
Coding vs. seeking/sharing knowledge

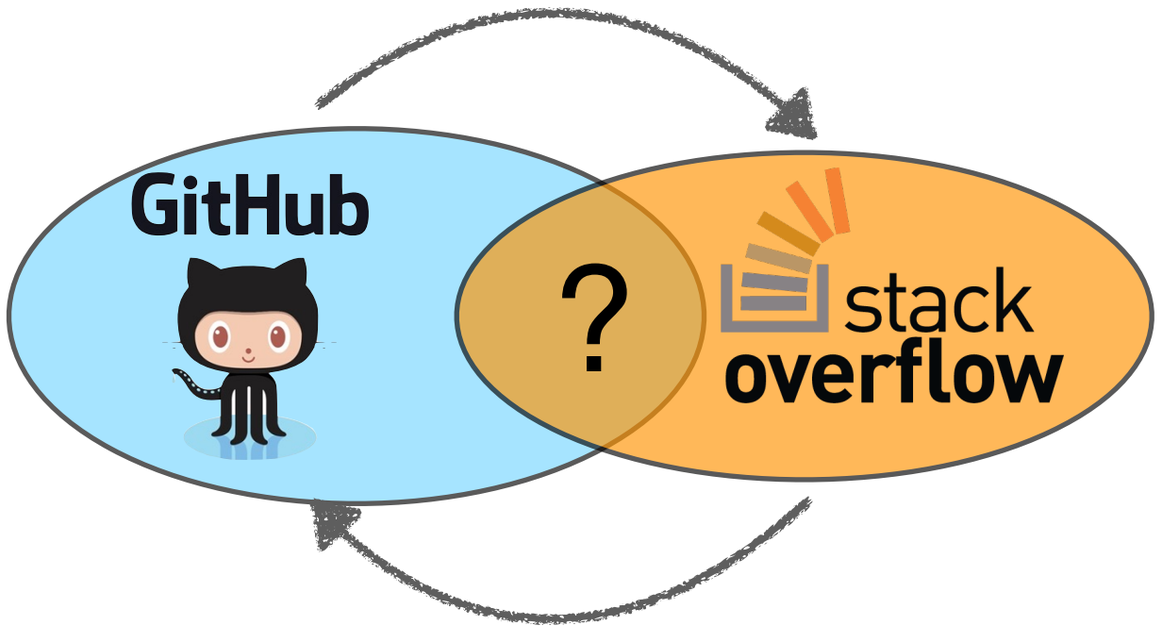




- Stack Overflow and GitHub: Associations between software development and crowdsourced knowledge
  B Vasilescu, V Filkov, A Serebrenik. *Social Computing 2013*

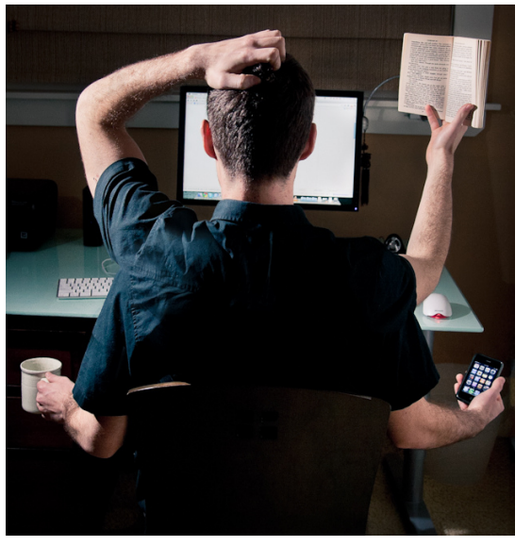
Coding vs. seeking/sharing knowledge

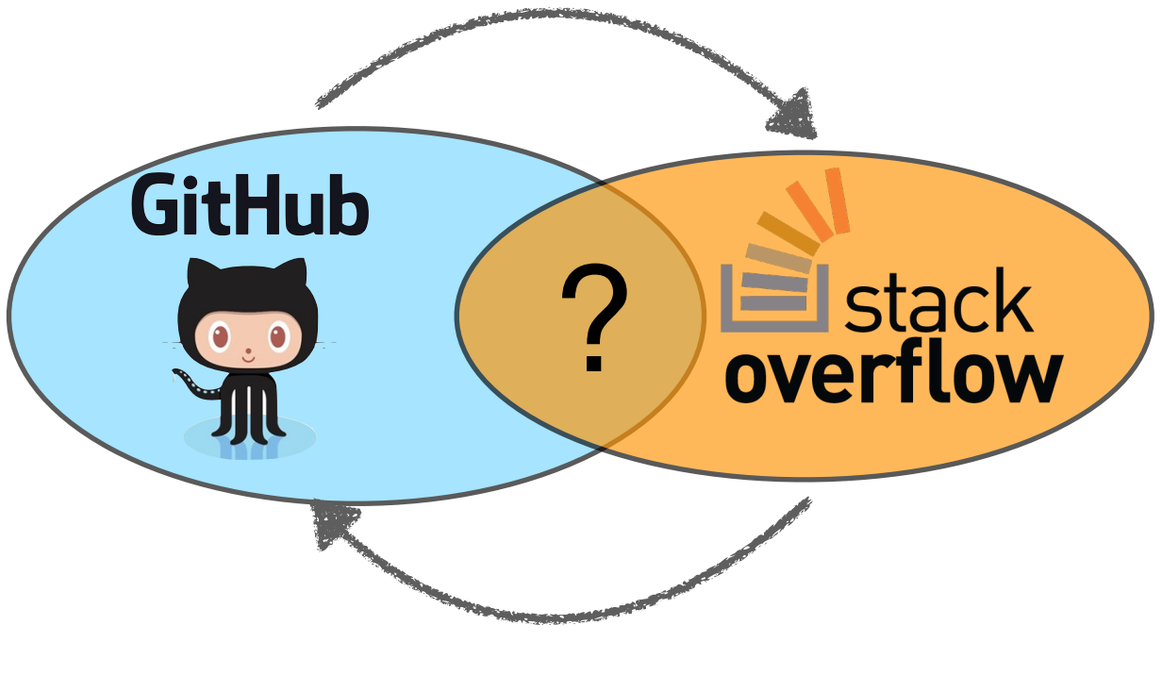


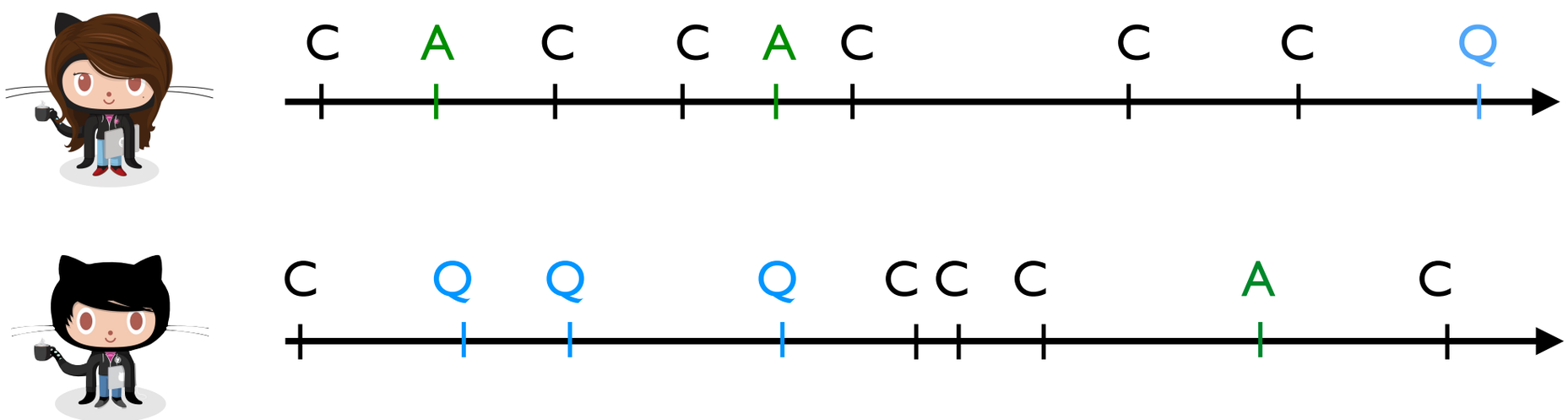


- Stack Overflow and GitHub: Associations between software development and crowdsourced knowledge
  B Vasilescu, V Filkov, A Serebrenik. *Social Computing 2013*

Coding vs. seeking/sharing knowledge

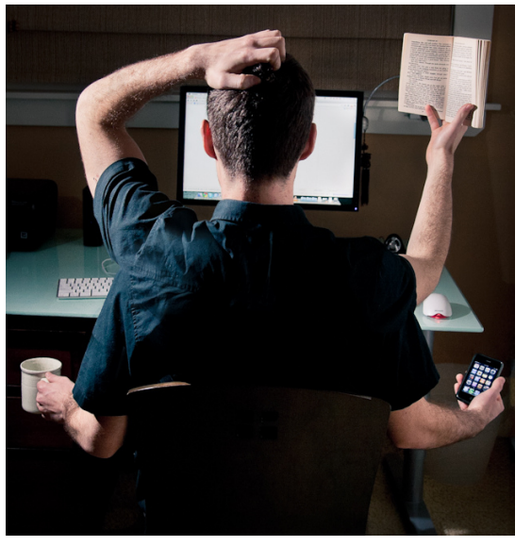


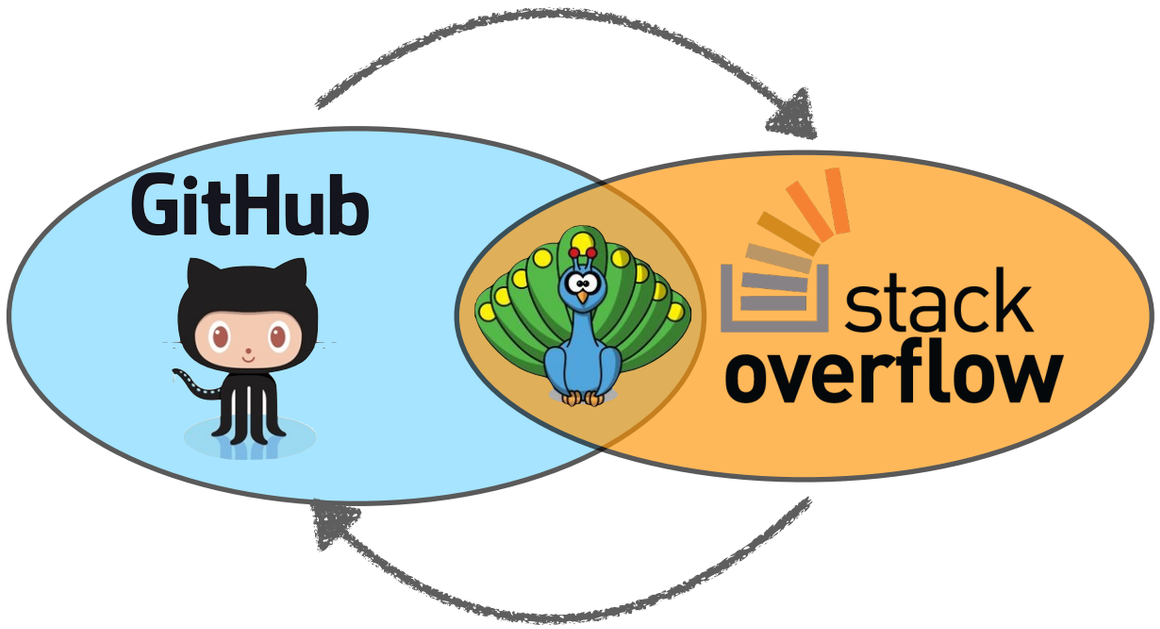




- Stack Overflow and GitHub: Associations between software development and crowdsourced knowledge
  B Vasilescu, V Filkov, A Serebrenik. *Social Computing 2013*

Coding vs. seeking/sharing knowledge



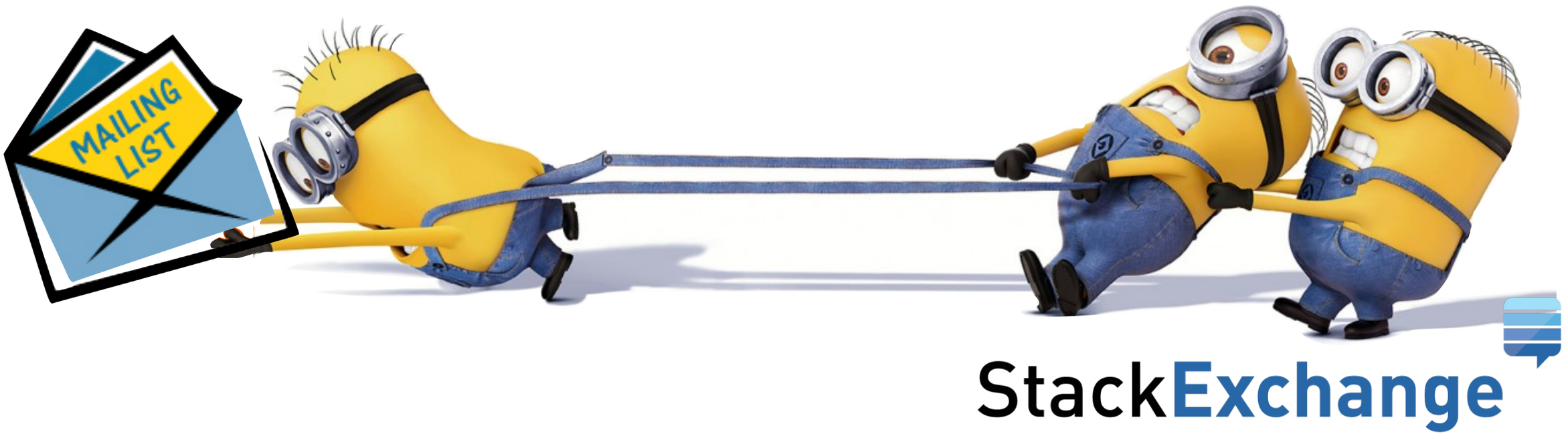


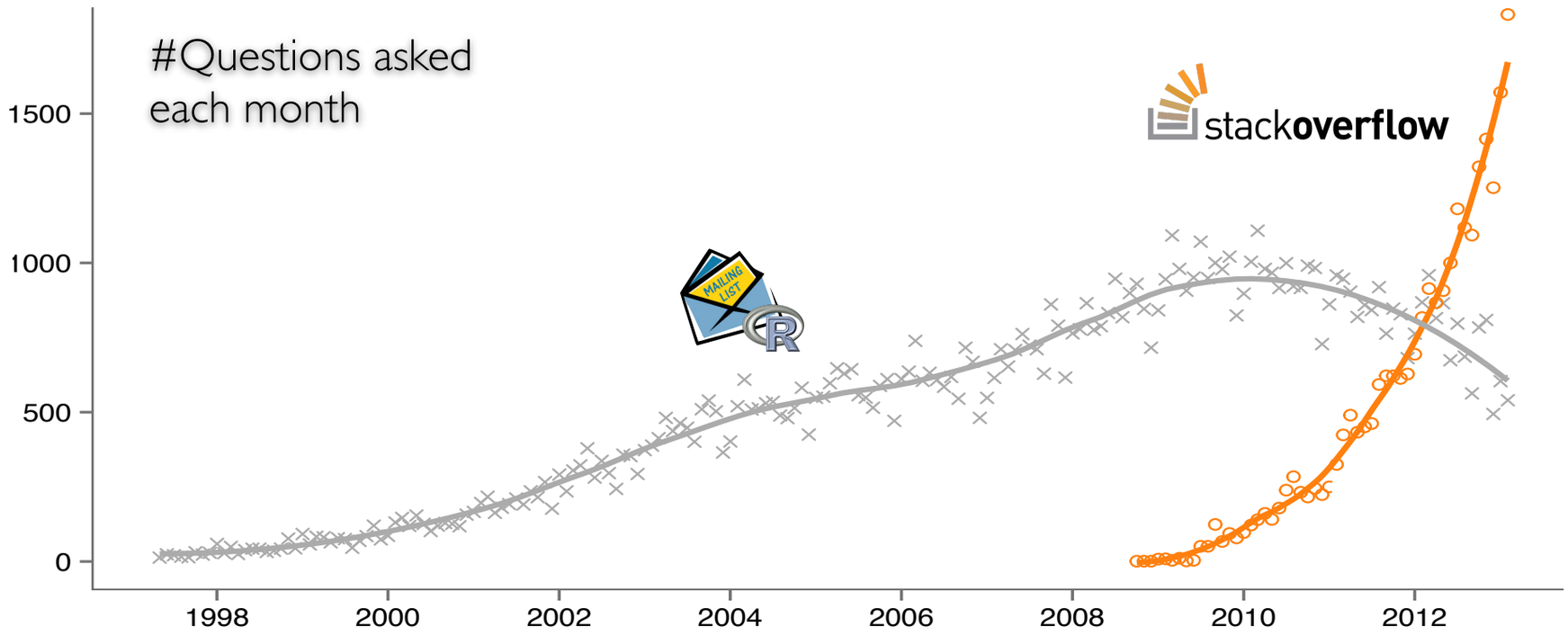
GitHub ~ stack overflow

- Experts are experts everywhere
- Participating in SO "speeds up" coding

- Stack Overflow and GitHub: Associations between software development and crowdsourced knowledge
  B Vasilescu, V Filkov, A Serebrenik. *Social Computing 2013*

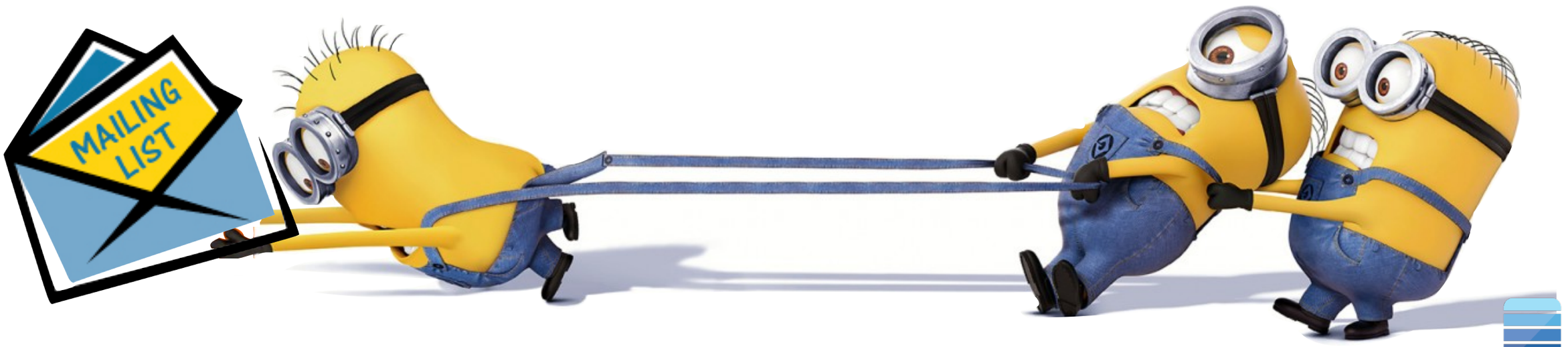
Gamification





- How social Q&A sites are changing knowledge sharing in open source software communities
  B Vasilescu, A Serebrenik, P Devanbu, V Filkov. *CSCW 2014*

Gamification



December 2003 Archives by thread

- **Messages sorted by:** [subject] [author] [date]
- **More info on this list...**

Starting: *Mon Dec 1 00:27:40 CEST 2003*
Ending: *Wed Dec 31 21:39:13 CEST 2003*
Messages: 1159

- [R] fitting a theoretical distribution with truncated tails *Piyush Sharma*
  - [R] fitting a theoretical distribution with truncated tails *Spencer Graves*
- [R] Discovering methods *Gabor Grothendieck*
  - [R] Discovering methods *Duncan Murdoch*
  - [R] Discovering methods *Thomas Lumley*
    - [R] Discovering methods *Duncan Murdoch*
      - [R] Discovering methods *Thomas Lumley*
- [R] hdf library for windows *Toby.Patterson at csiro.au*
- [R] hdf library for windows *Mulholland, Tom*

stackoverflow Questions Tags Users Badges Unanswered

Tagged Questions info newest 2 featured frequent votes active unanswered

R is a free, open source programming language and software environment for statistical computing, bioinformatics and graphics. It is advised to supplement your question with a minimal reproducible example. For statistical questions please use stats.stackexchange.com.

learn more... | improve tag wiki | top users | synonyms (1)

431 votes
11 answers
24k views

How to make a great R reproducible example?

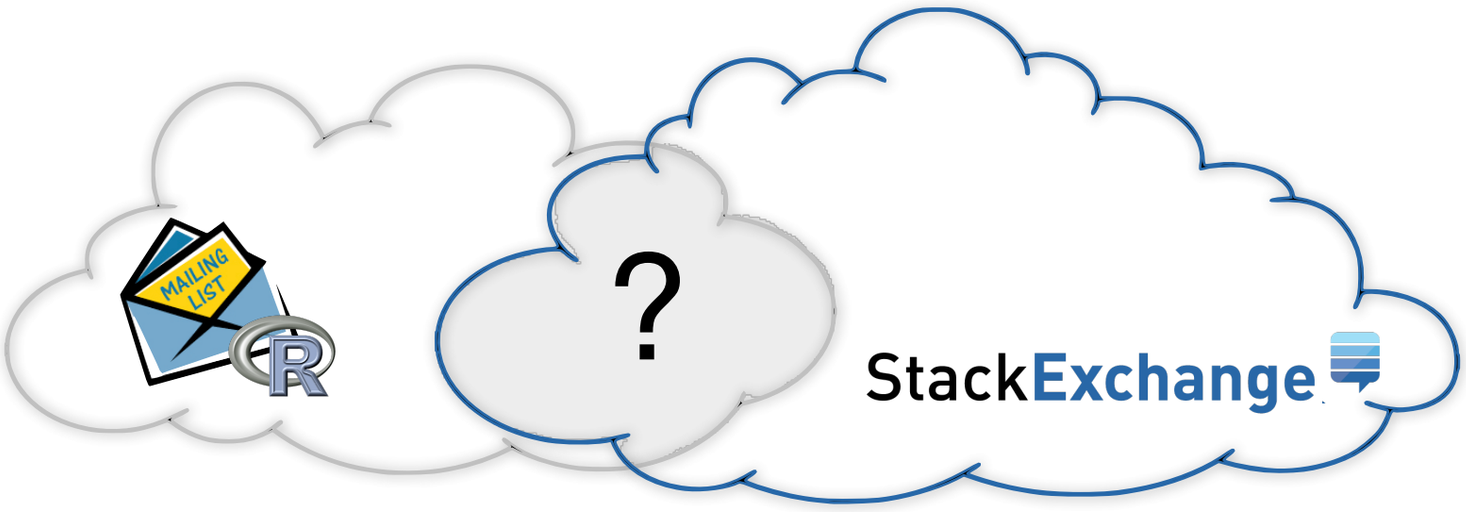
When discussing performance with colleagues, teaching, sending a bug report or searching for guidance on mailing lists and here on SO, a reproducible example is often asked and always helpful. What ...

r r-faq

community wiki
4 revs, 3 users 62%
Andrie

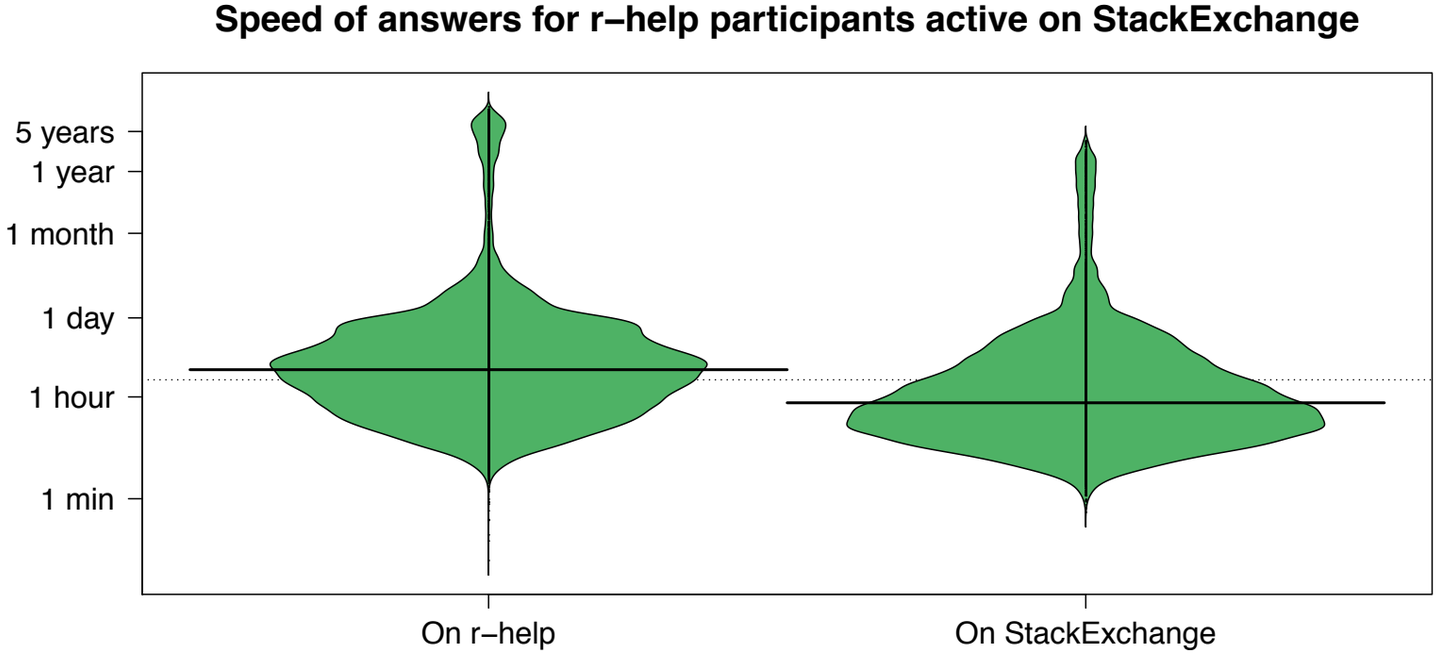
- How social Q&A sites are changing knowledge sharing in open source software communities
  B Vasilescu, A Serebrenik, P Devanbu, V Filkov. *CSCW 2014*

Gamification



- How social Q&A sites are changing knowledge sharing in open source software communities
  B Vasilescu, A Serebrenik, P Devanbu, V Filkov. *CSCW 2014*

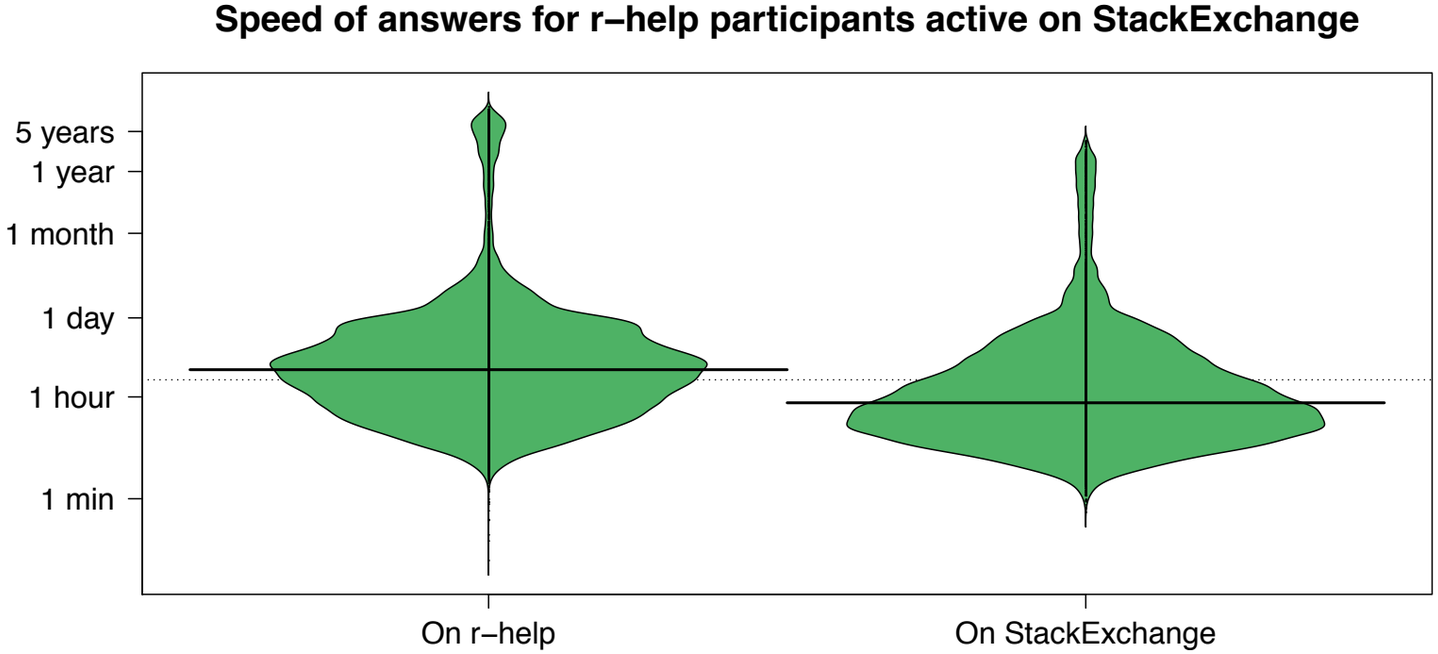
Gamification



median ~3h median 47mins

- How social Q&A sites are changing knowledge sharing in open source software communities
  B Vasilescu, A Serebrenik, P Devanbu, V Filkov. *CSCW 2014*

Gamification



median ~3h

median 47mins

"The reputation ratings are a nice little incentive"

"It's a game, which also serves a good purpose"

"Peer recognition/ gamification within Stack Overflow"

- How social Q&A sites are changing knowledge sharing in open source software communities
  B Vasilescu, A Serebrenik, P Devanbu, V Filkov. *CSCW 2014*

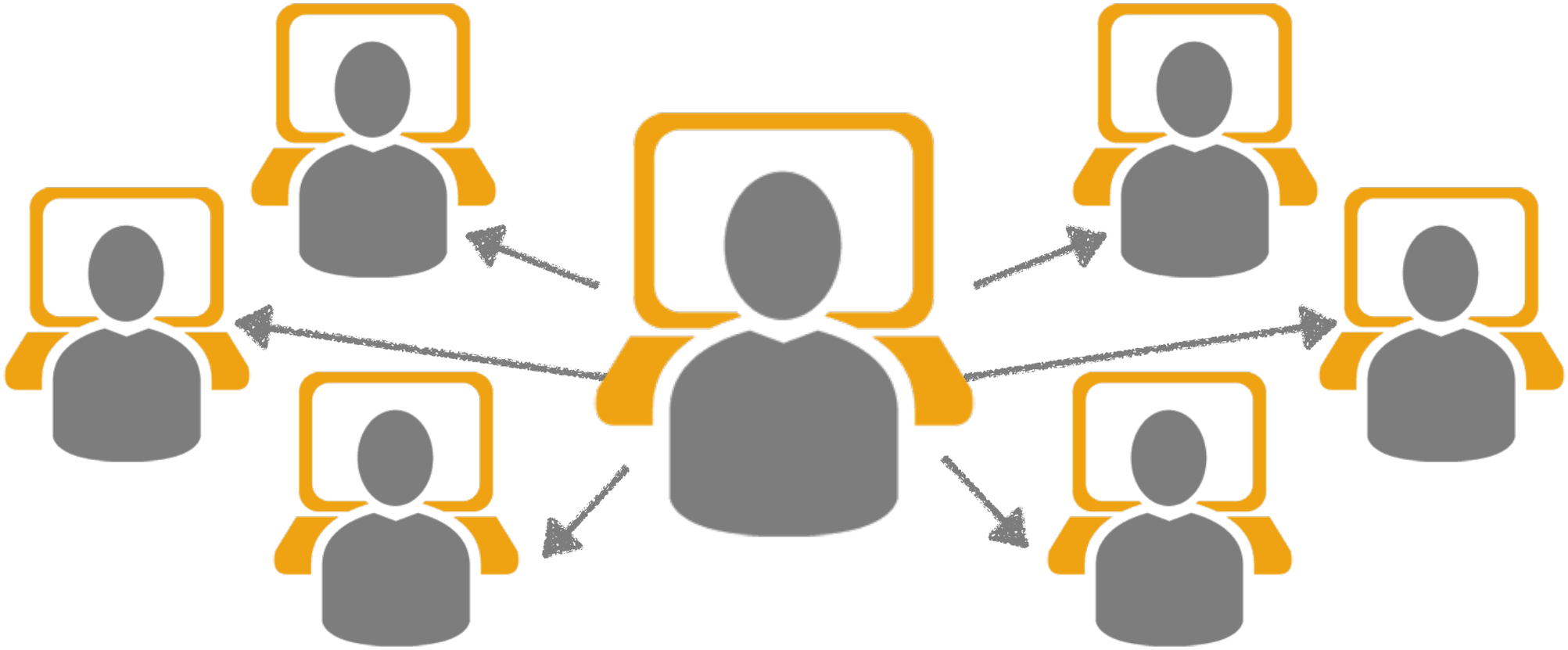
The “social programmer”

… has a public profile

… shares code

… is social

… signals



t day

t day

1

The social cost of "social coding"

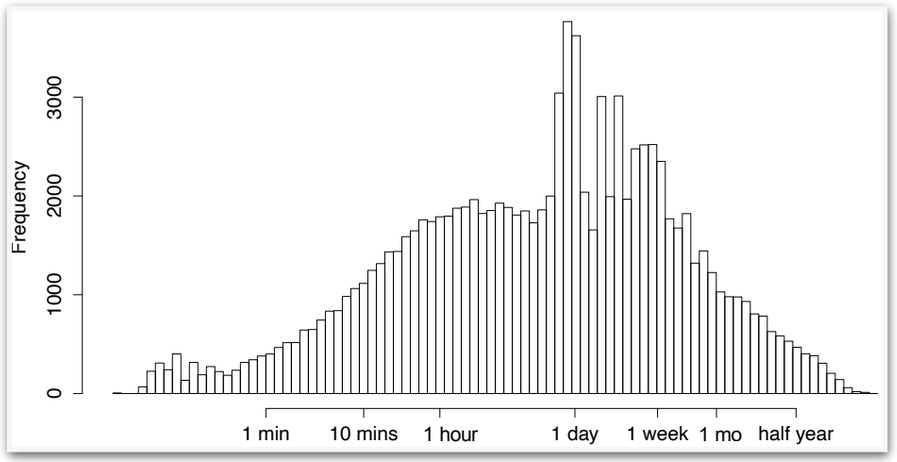
t day

1

2

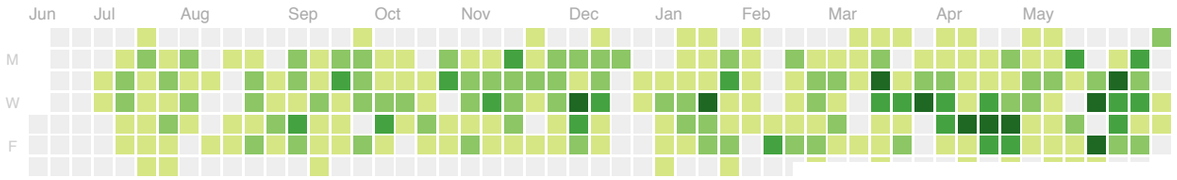
The social cost of "social coding"

The predictability of "social coding"



1 The social cost of “social coding”

Public contributions



Summary of Pull Requests, issues opened, and commits. Learn more.

Contributions in the last year	Longest streak
1,365 total	12 days
Jun 19, 2014 – Jun 19, 2015	April 27 – May 8

Most active GitHub users (git.io/top (http://git.io/top))

#	User	Contribs	Language	
#1	Ocramius (https://github.com/Ocramius) (Marco Pivetta)	61259	PHP	Frankfurt
#2	michalbe (https://github.com/michalbe) (Michał Budzyński)	14771	JavaScript	Warsaw, Poland
#3	kevinsawicki (https://github.com/kevinsawicki) (Kevin Sawicki)	9159	JavaScript	Redwood City, C

Jon Skeet
360 5367 6604
member for: 6 years, 8 months
#1 all time rank
784,585 all time reputation

Darin Dimitrov
98 2027 2062
member for: 6 years, 8 months
#2 all time rank
595,006 all time reputation

BalusC
142 1872 2174
member for: 5 years, 10 months
#3 all time rank
573,150 all time reputation

Hans Passant
61 656 1236
member for: 6 years, 9 months
#4 all time rank
553,678 all time reputation

Women are less effective than men in mixed-gender competitive environments [Gneezy et al]

Women shy away from competition and men embrace it [Niederle and Vesterlund]

- Performance in competitive environments: Gender differences
  U Gneezy, M Niederle, A Rustichini. *The Quarterly Journal of Economics, 118(3):1049–1074, 2003*
- Do women shy away from competition? Do men compete too much?
  M Niederle, L Vesterlund. *The Quarterly Journal of Economics, 122(3):1067–1101, 2007*

1

Gender representation

stackoverflow

5.8%

- Stack Overflow 2015 Developer Survey (26,086 people from 157 countries)
  http://stackoverflow.com/research/developer-survey-2015#profile-gender

github SOCIAL CODING

~5-9%

- Exploring the data on gender and GitHub repo ownership
  Alyssa Frazee. http://alyssafrazee.com/gender-and-github-code.html
- Gender and tenure diversity in GitHub teams
  B Vasilescu, D Posnett, B Ray, M vd Brand, A Serebrenik, P Devanbu, V Filkov. *CHI 2015*

1

Gender representation

stackoverflow

5.8%

- Stack Overflow 2015 Developer Survey (26,086 people from 157 countries)
  http://stackoverflow.com/research/developer-survey-2015#profile-gender

github SOCIAL CODING

~5-9%

- Exploring the data on gender and GitHub repo ownership
  Alyssa Frazee. http://alyssafrazee.com/gender-and-github-code.html
- Gender and tenure diversity in GitHub teams
  B Vasilescu, D Posnett, B Ray, M vd Brand, A Serebrenik, P Devanbu, V Filkov. *CHI 2015*

open source™

10.9%

Google

18%

16.6%

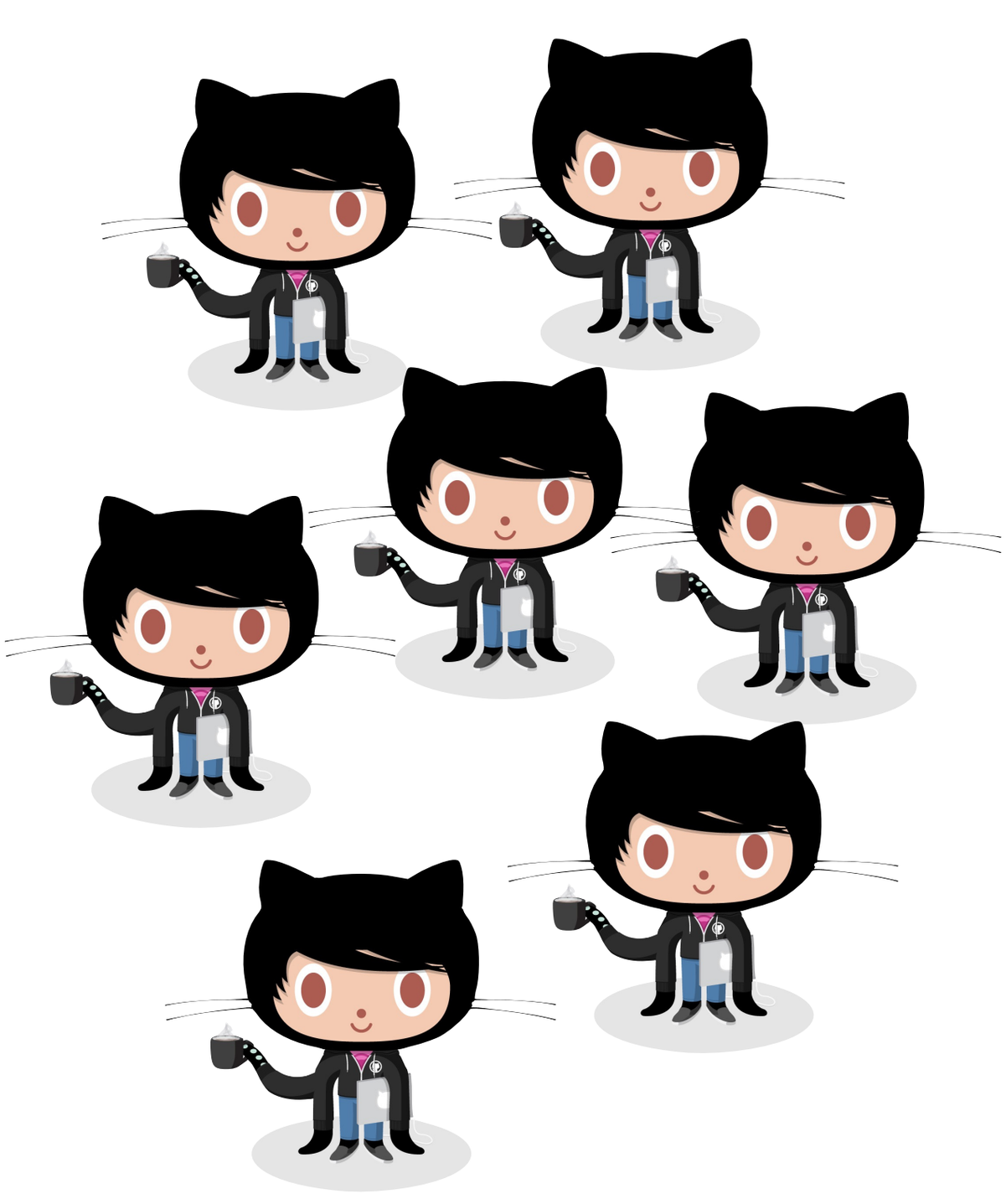
- FLOSS 2013: A survey dataset about free software contributors: challenges for curating, sharing, and combining
  G Robles, L Arjona-Reina, B Vasilescu, A Serebrenik, JM Gonzalez-Barahona. *MSR 2014*
- Google Diversity (2015) www.google.com/diversity/index.html#chart
- Inside Microsoft (2015) https://www.microsoft.com/en-us/diversity/inside-microsoft/default.aspx#fbid=cynGfV-jQoE

Which is more effective?

- Gender and tenure diversity in GitHub teams
  B Vasilescu, D Posnett, B Ray, M vd Brand, A Serebrenik, P Devanbu, V Filkov. *CHI 2015*

1

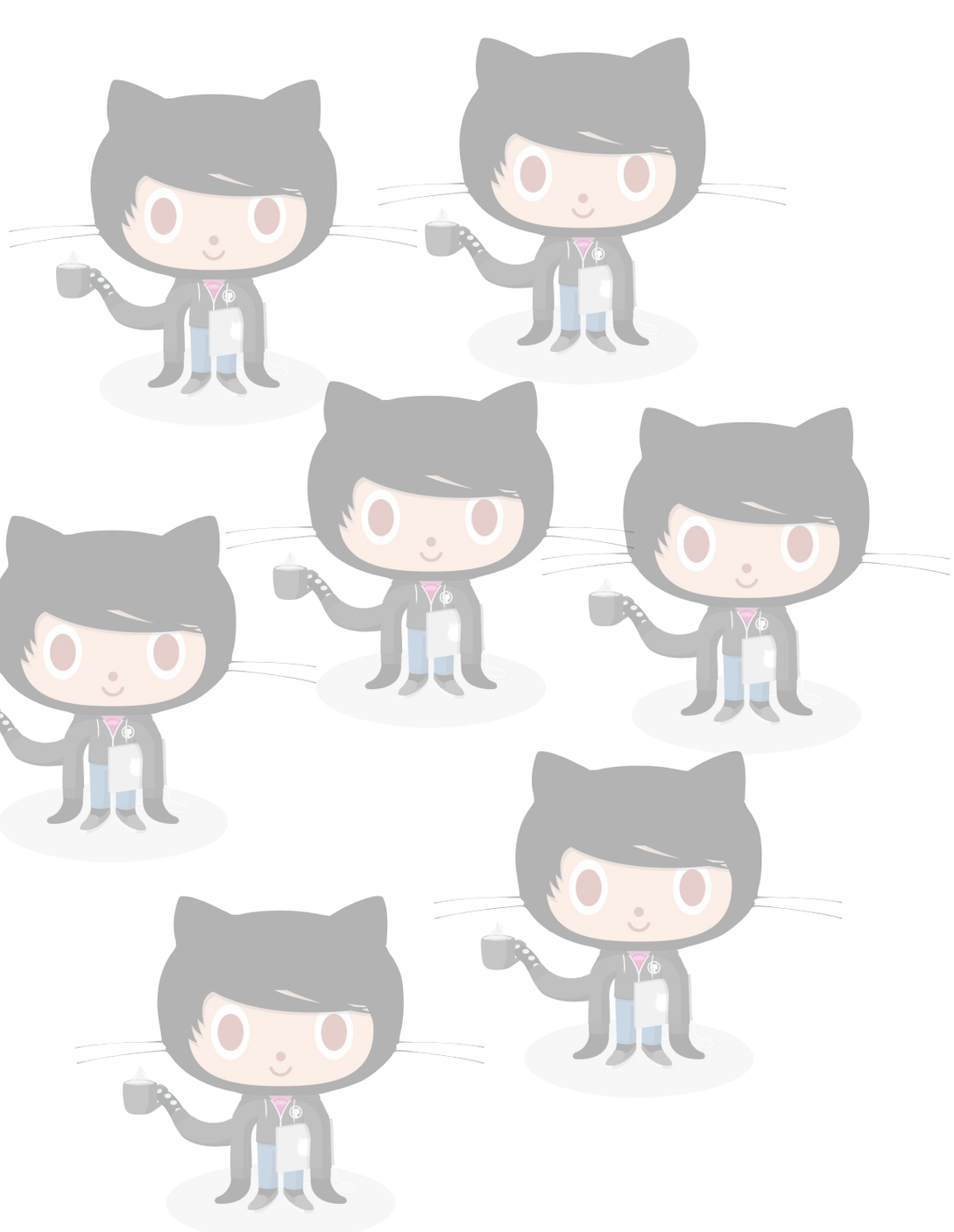
Which is more effective?



- Gender and tenure diversity in GitHub teams
  B Vasilescu, D Posnett, B Ray, M vd Brand, A Serebrenik, P Devanbu, V Filkov. *CHI 2015*

1

Which is more effective?



- Gender and tenure diversity in GitHub teams
  B Vasilescu, D Posnett, B Ray, M vd Brand, A Serebrenik, P Devanbu, V Filkov. *CHI 2015*

1

Diversity 👎

Similarity attraction theory

People prefer working with others similar to them in terms of values, beliefs, and attitudes [Byrne]

Social identity and social categorization theory

People categorize themselves into specific groups. Members of own group are treated better than outsiders [Tajfel]

Due to greater perceived differences between groups than within groups, diversity can lead to confusion, stress, and conflict [Horwitz & Horwitz]

Diversity 👍

Driver of internal innovation and **business growth** [Forbes]

Diverse problem solvers **outperform high ability problem solvers** [Hong & Page]

Companies with diverse executive boards have higher earnings and returns on equity [McKinsey]

Multicultural social networks promote creativity [Harvard Business School]

1

Diversity 👍

Information Processing Theory

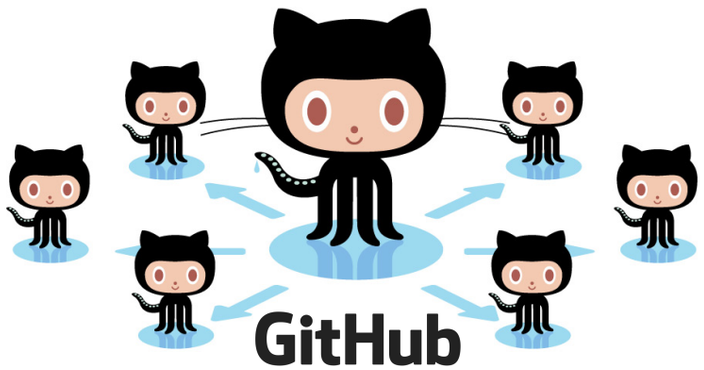
Mixture of cultural/educational backgrounds
+ access to different networks/broader information
=> creativity, adaptability, & problem solving skills.
[Salancik & Pfeffer]

1

Today: diversity in open source software (OSS) GitHub teams

Different settings

Geographic & cultural dispersion



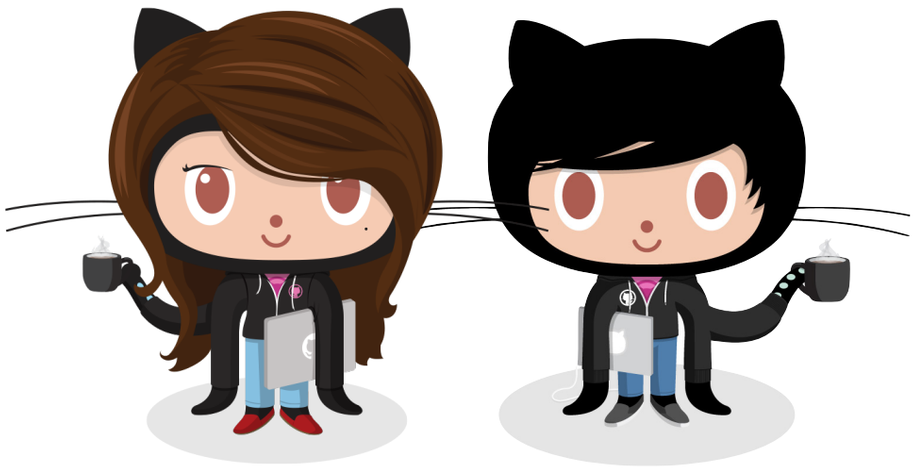
Online communities & distributed comm. channels

Different methods

Quantitative; large-scale trace data

1

Today: gender & tenure diversity in open source software (OSS) GitHub teams



Gender diversity
= mix women/men

simplifying assumption: gender is binary

The "hacker" culture is male-dominated and unfriendly to women [Turkle]

Women are <10% in OSS [Robles et al]

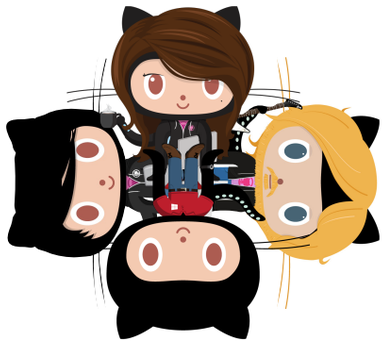
Reports of active discrimination and sexism towards women [Nafus]

1

Today: gender & tenure diversity in open source software (OSS) GitHub teams

Tenure diversity
= mix junior/senior

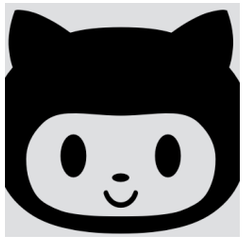
The "onion" structure of OSS: small (stable) core + large (loose) periphery [Ducheneaut]



High turnover [Robles & Gonzalez-Barahona]

1

Today: gender & tenure diversity in open source software (OSS) GitHub teams



World's largest open source community

Trace data available @ghtorrent
[Gousios et al]

1

Today: gender & tenure diversity in open source software (OSS) GitHub teams

Theoretical

Technical

OSS as meritocracy; contribution quality as main driver of impression formation
[Dabbish et al, Marlow et al]

1

Today: gender & tenure diversity in open source software (OSS) GitHub teams

Theoretical

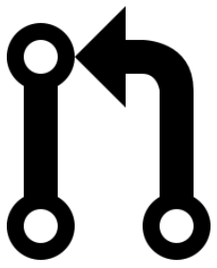
Technical

Demographics are less salient in OSS [Riordan & Shore]

1

Today: gender & tenure diversity in open source software (OSS) GitHub teams

Theoretical



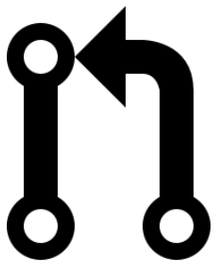
Technical

Anyone can contribute to any repository.
Who's on a team?

1

Today: gender & tenure diversity in open source software (OSS) GitHub teams

Theoretical



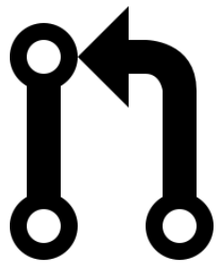
Technical

Gender is not explicitly recorded

1

Today: gender & tenure diversity in open source software (OSS) GitHub teams

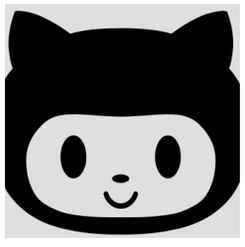
Theoretical



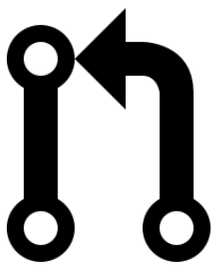
Technical

People contribute under multiple aliases

Today: gender & tenure diversity in open source software (OSS) GitHub teams



Theoretical

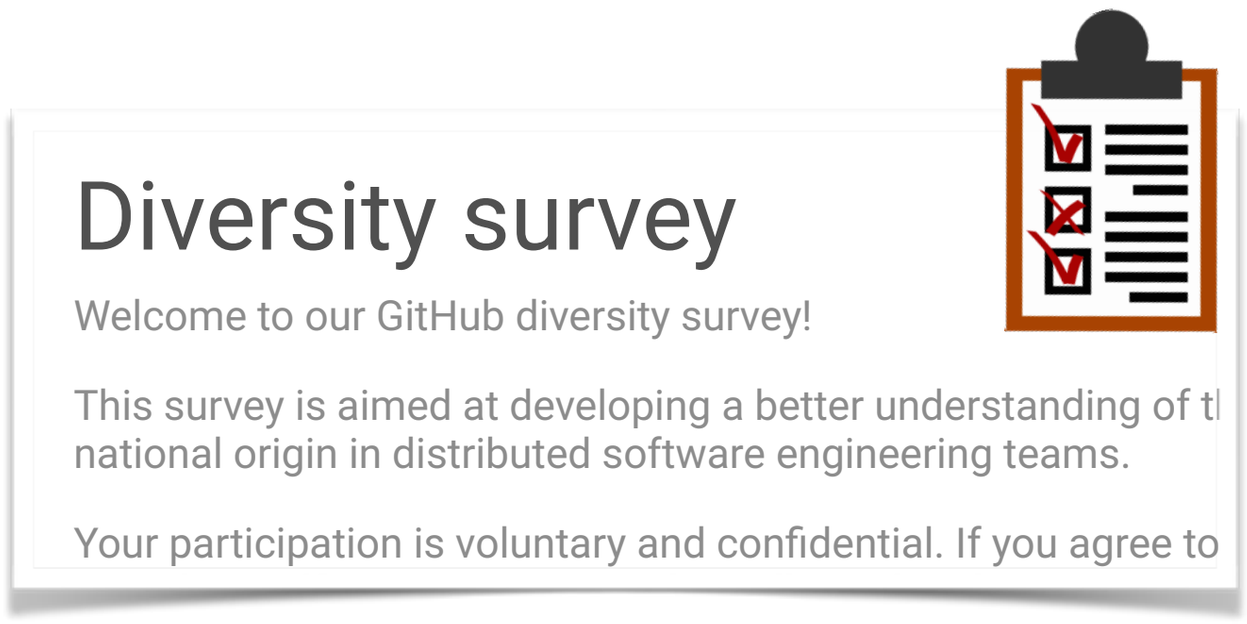


Technical

How to analyze such large-scale longitudinal trace data?

1

Approach: mixed methods

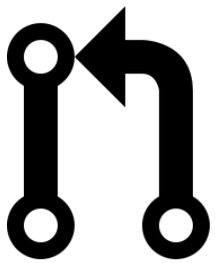


+

- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

Survey

4,500 invitations, 816 responses



What constitutes a team?

Which differences do people recognize among team members?

Does diversity matter?

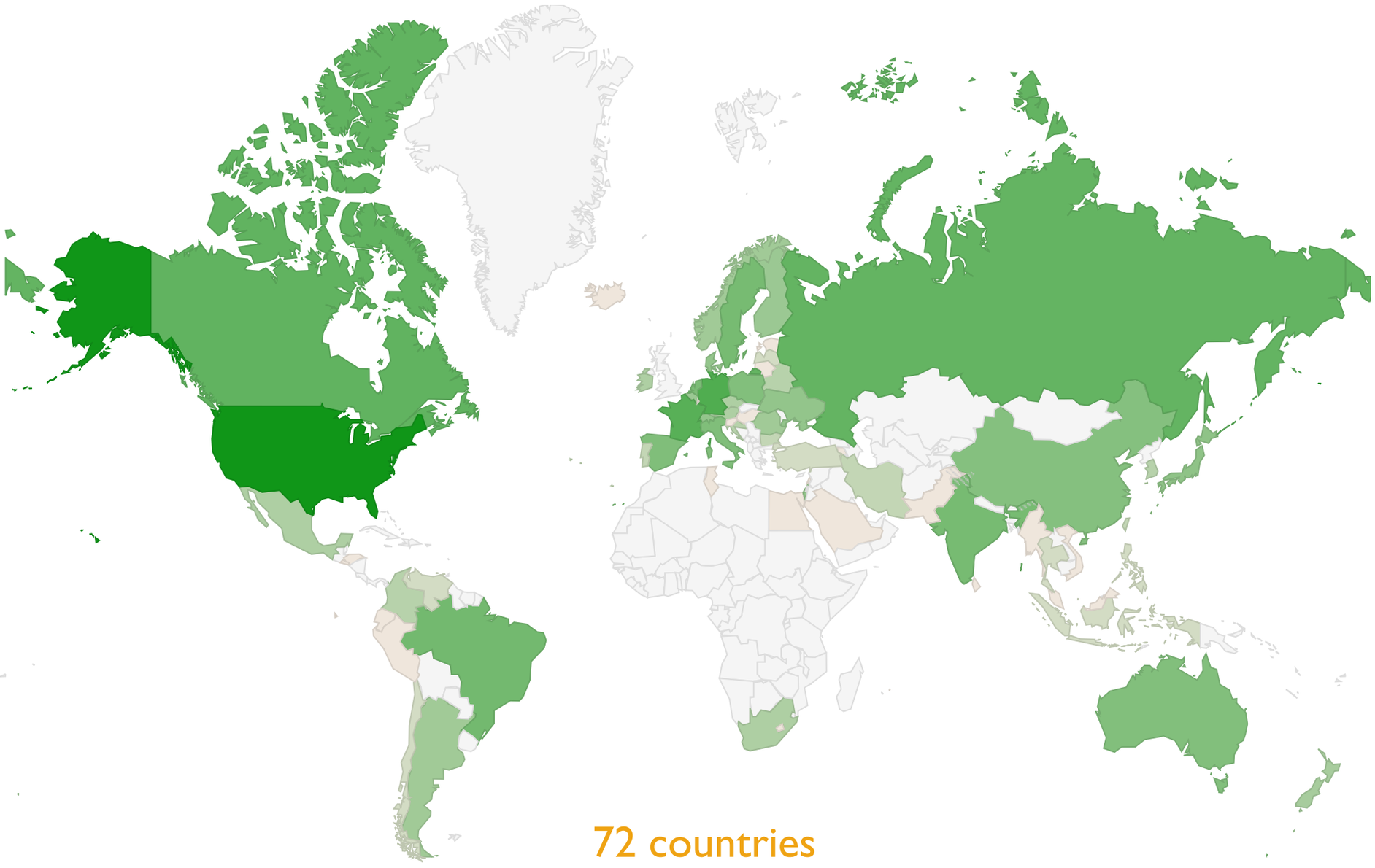
- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

Survey

4,500 invitations, 816 responses

F 24%
M 75%



72 countries

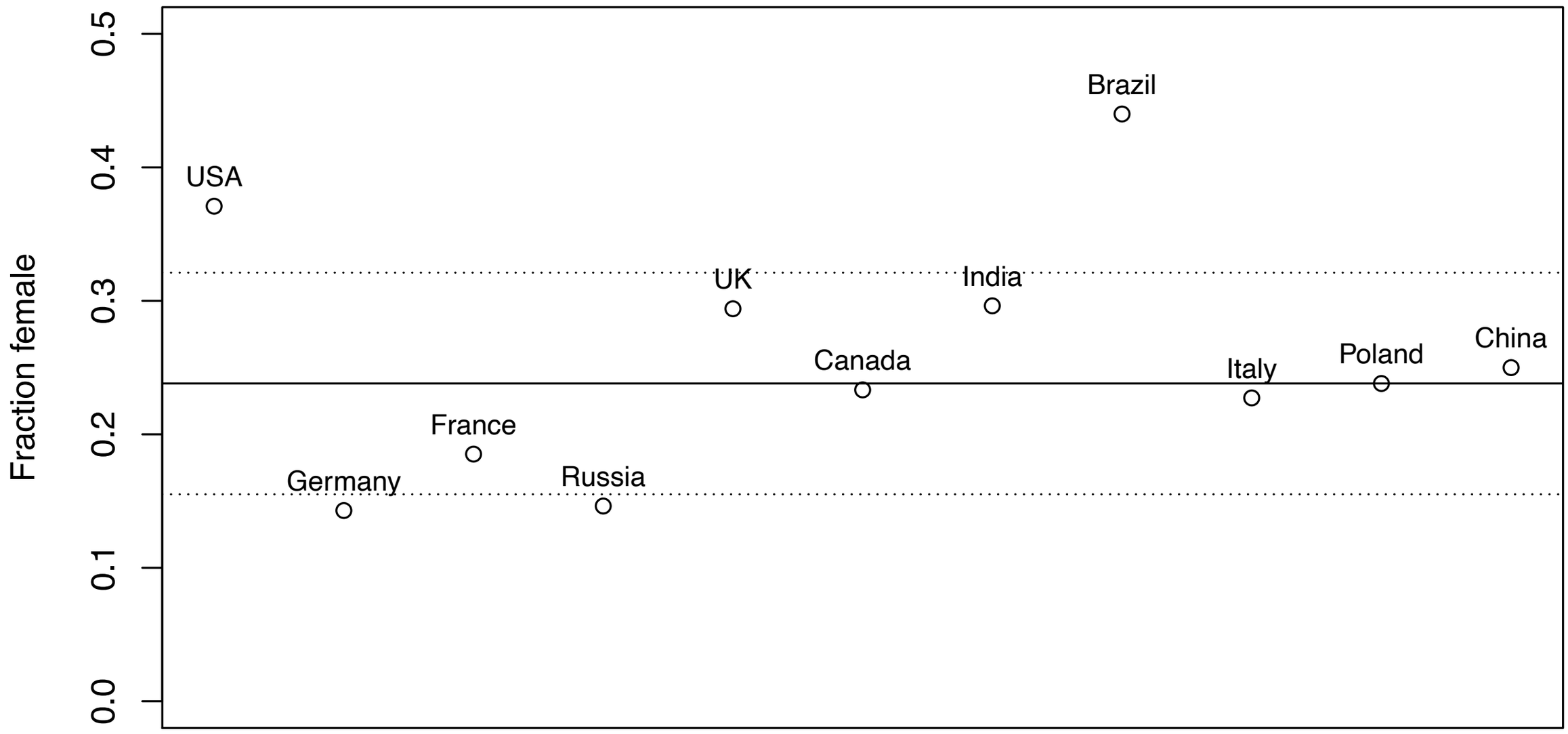
- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

Survey

4,500 invitations, 816 responses

F 24%
M 75%



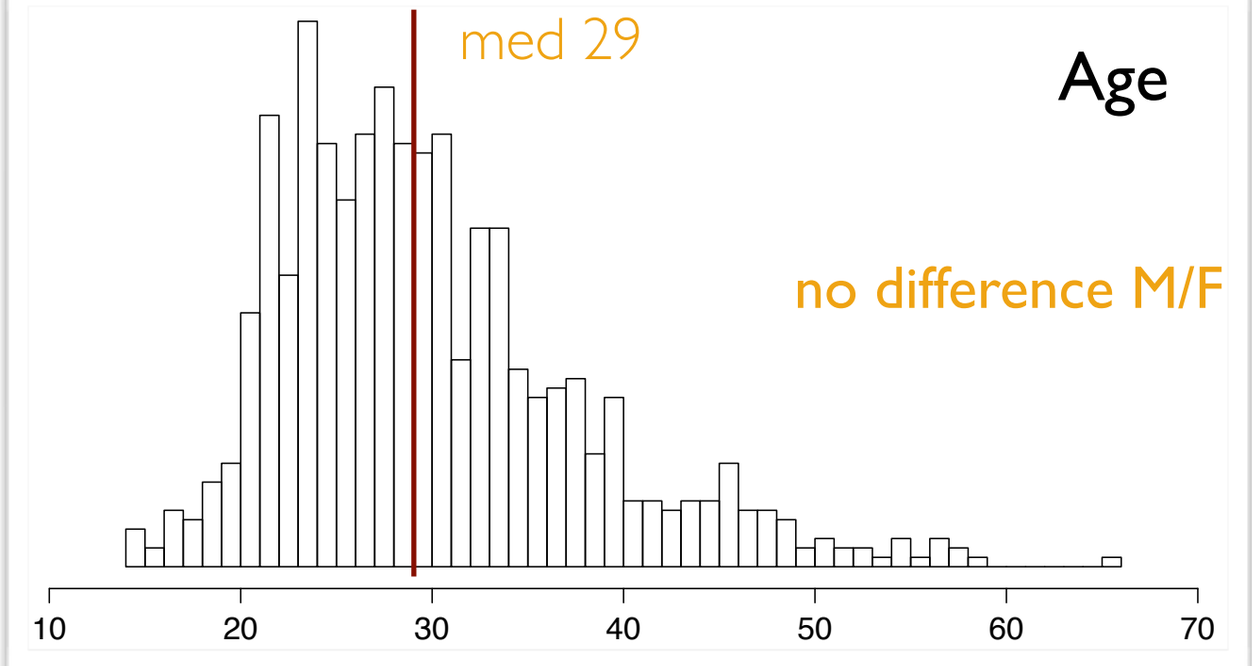
- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

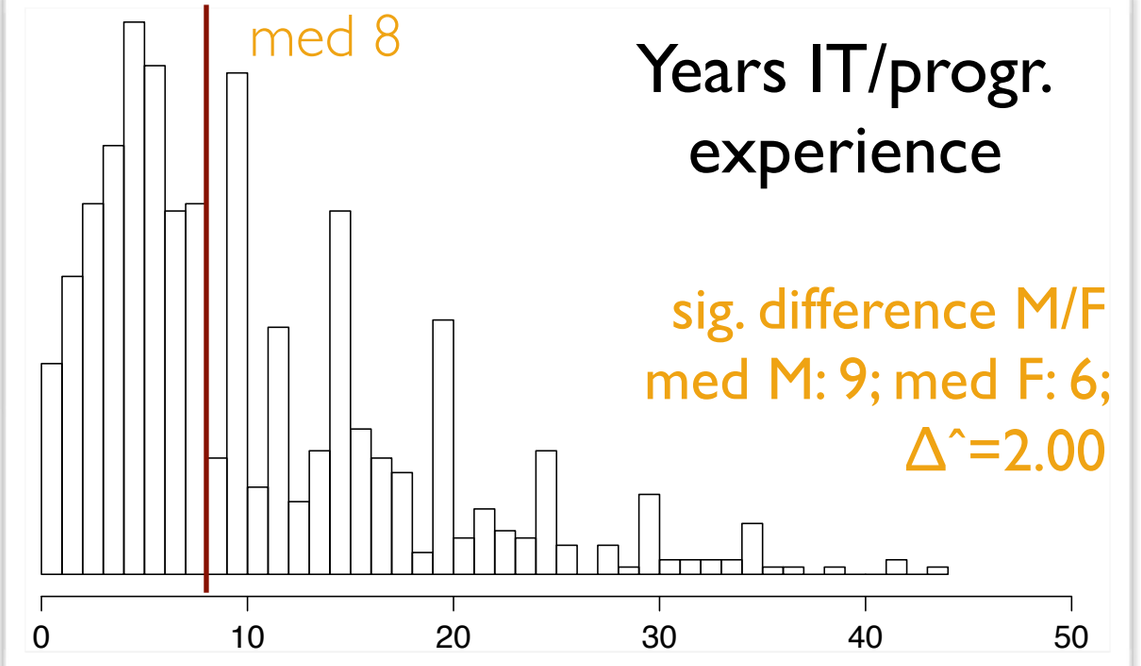
1

Survey

4,500 invitations, 816 responses

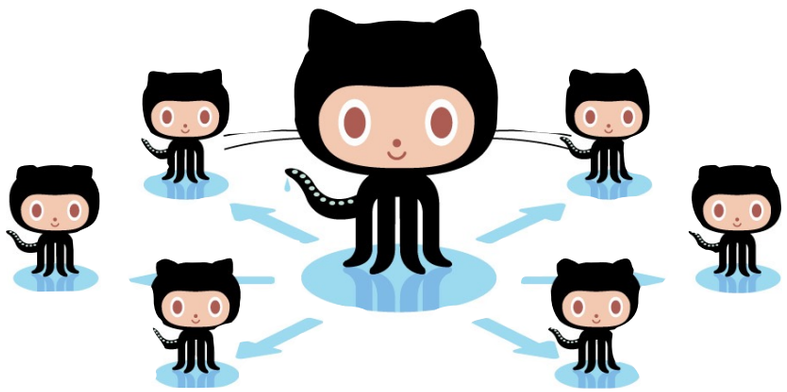
F 24%
M 75%





- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

What constitutes a team?



- The repository owner and others who can push directly
- People who contribute code frequently
- People who work on my particular feature/ branch

less inclusive

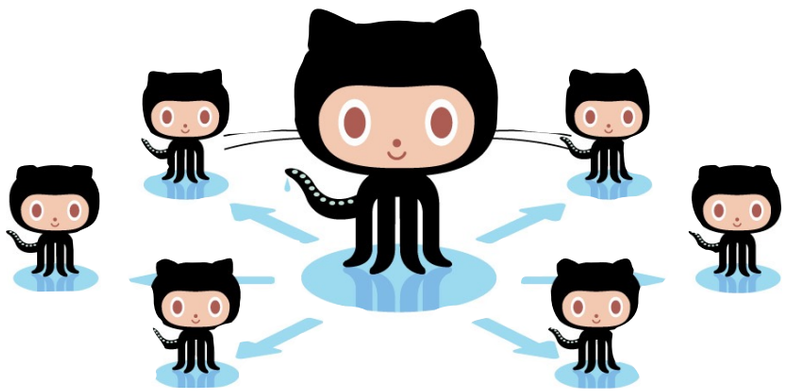
more inclusive

- Everyone who does something in this repository

- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

What constitutes a team?



- The repository owner and others who can push directly
- People who contribute code frequently
- People who work on my particular feature/ branch

less inclusive

more inclusive

#1 (72%)

- Everyone who does something in this repository

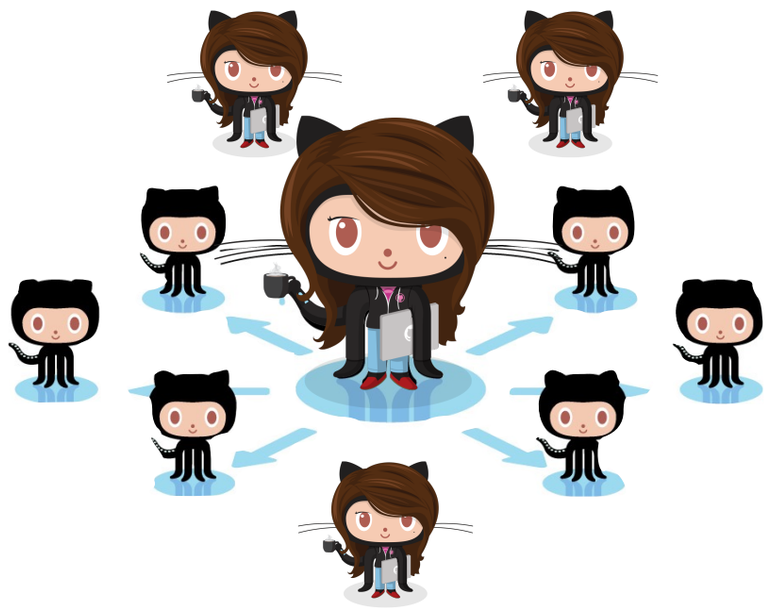
Everyone

- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

Which differences do people recognize among team members?

… for (none other / few other / most other) team members



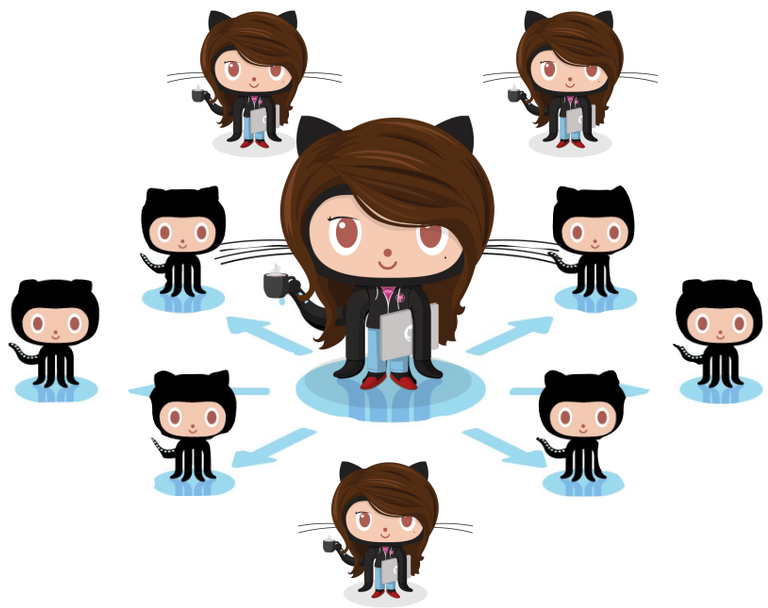
- Programming skills
- Social skills
- Gender
- Ethnicity
- Overall GitHub experience
- Reputation as programmer
- Country of residence
- Personality
- Age
- Educational level
- Real name
- Hobbies
- Employment
- Political views

- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

Which differences do people recognize among team members?

… for (none other / few other / **most other**) team members



Demographics not salient is OSS [Riordan & Shore]

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

Developers are aware of each other's gender

- **Perceptions of diversity on GitHub: A user survey**
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

Does diversity matter?

"code sees no color or gender"

"any **demographic identity is irrelevant**"

"**more about the contributions to the code** than the 'characteristics' of the person"

Meritocracy; no effects of diversity

- **Perceptions of diversity on GitHub: A user survey**
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

Does diversity matter?

"diverse viewpoints often lead to **lively discussions and new ideas**"

"in general it is always **enriching** to communicate with someone different"

"diversity in the body of folks willing to interact and contribute works to **strengthen the usability** of the library"

Positive effects of diversity

- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

Does diversity matter?

Gender related

"I have used a fake GitHub handle (my normal GitHub handle is my first name, which is a distinctly female name) **so that people would assume I was male**"

"interactions are usually positive too, with occasional sexism, but **nothing more then one encounters in the rest of life**"

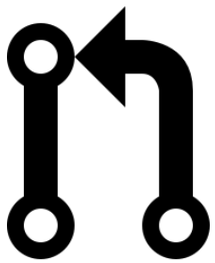
"… caused me to **leave a project**"

Negative effects of diversity

- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

Survey

4,500 invitations, 816 responses



What constitutes a team?

The team is everyone

Which differences do people recognize among team members?

Gender is surprisingly salient

Does diversity matter?

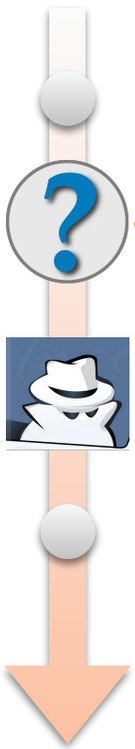
Positive/negative/no effects of diversity

- Perceptions of diversity on GitHub: A user survey
  B Vasilescu, V Filkov, A Serebrenik. *CHASE 2015*

1

Mining

github SOCIAL CODING



Infer genders
[Vasilescu et al, IWC'13]

Sample
4K projects

Bogdan Vasilescu
bvasiles

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

Contributions | Repositories | Publi

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

Contributions

Apr May Jun Jul Aug Sep

Bing Maps + Heuristics
http://github.com/tue-mdse/countryNameManager

Bogdan + USA

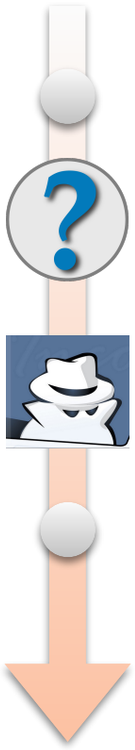
Name frequency tables for 30 countries
http://github.com/tue-mdse/genderComputer

male

Andrea + Italy = male
Andrea + USA = female

1

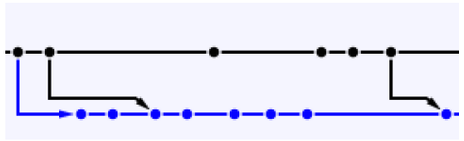
Mining

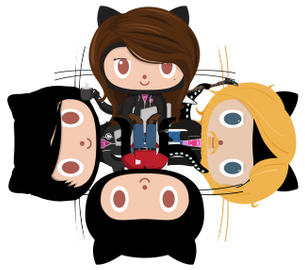


Sample
4K projects

Response

Productivity
(#commits/quarter)





Turnover
(fraction team new
w.r.t. prev. quarter)

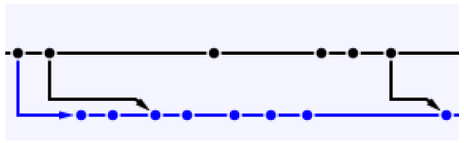
1

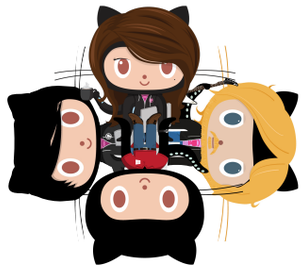
Mining

Sample
4K projects

Response

Productivity
(#commits/quarter)





Turnover
(fraction team new
w.r.t. prev. quarter)

Independent

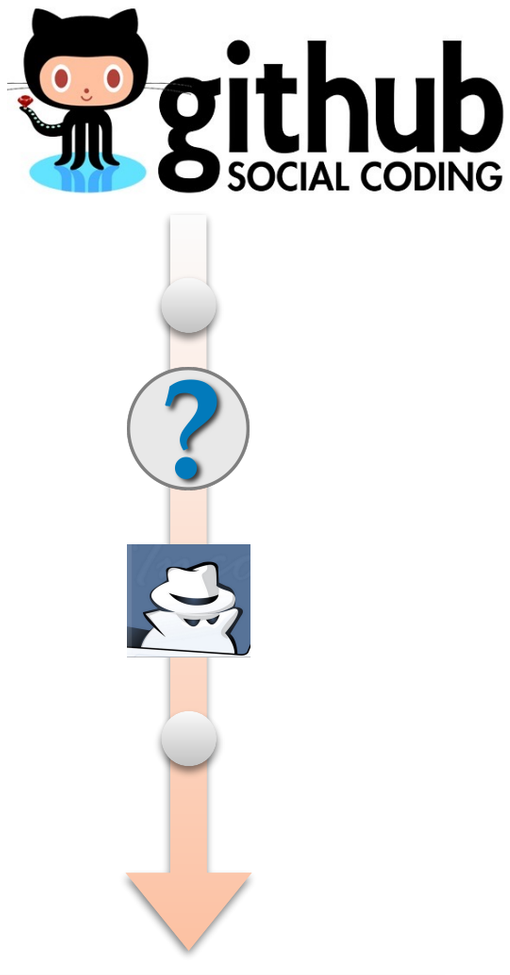
Gender
diversity
(Blau index)

Tenure diversity
(coeff. variation)

- project
- overall coding

1

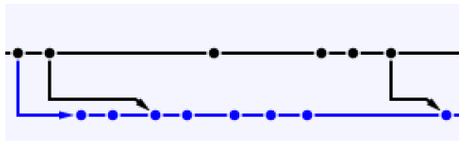
Mining

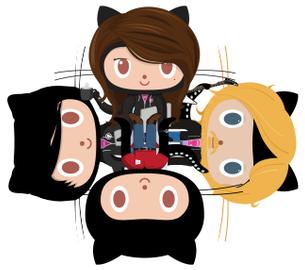


Sample
4K projects

Response

Productivity
(#commits/quarter)





Turnover
(fraction team new
w.r.t. prev. quarter)

Independent

Gender
diversity
(Blau index)

Tenure diversity
(coeff. variation)

- project
- overall coding

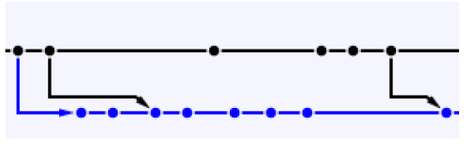
Controls

Team size

Time

Project age

Project activity



…

Analysis

Nesting: projects

Project	Created on	Project age	Total #commits	#Forks	Time	#Commits	#Comments	Team size	Gender diversity	Commit tenure diversity	Turnover
A	2011-02-15	12	557	51	Q2	47	26	9	0.25	0.47	0.67
					Q5	19	12	10	0.00	0.93	0.75
					Q6	7	13	12	0.25	0.54	0.67
					Q7	56	53	20	0.00	0.56	0.87
					…						
B	2010-09-21	11	2075	578	Q4	71	169	83	0.03	0.66	0.87
					Q5	116	219	93	0.05	0.73	0.56
					Q6	186	367	119	0.06	0.80	0.86
					Q7	129	453	114	0.08	0.85	0.82
					…						

Analysis

Nesting: projects
Cross-classification: quarters

Project	Created on	Project age	Total #commits	#Forks	Time	#Commits	#Comments	Team size	Gender diversity	Commit tenure diversity	Turnover
A	2011-02-15	12	557	51	Q2	47	26	9	0.25	0.47	0.67
					Q5	19	12	10	0.00	0.93	0.75
					Q6	7	13	12	0.25	0.54	0.67
					Q7	56	53	20	0.00	0.56	0.87
					…						
B	2010-09-21	11	2075	578	Q4	71	169	83	0.03	0.66	0.87
					Q5	116	219	93	0.05	0.73	0.56
					Q6	186	367	119	0.06	0.80	0.86
					Q7	129	453	114	0.08	0.85	0.82
					…						

Analysis

Nesting: projects
Cross-classification: quarters

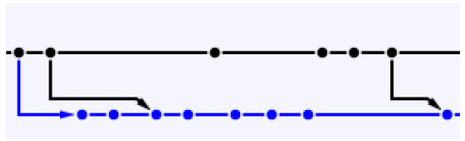
Linear mixed-effects (hierarchical) models

Project	Created on	Project age	Total #commits	#Forks	Time	#Commits	#Comments	Team size	Gender diversity	Commit tenure diversity	Turnover
A	2011-02-15	12	557	51	Q2	47	26	9	0.25	0.47	0.67
					Q5	19	12	10	0.00	0.93	0.75
					Q6	7	13	12	0.25	0.54	0.67
					Q7	56	53	20	0.00	0.56	0.87
					…						
B	2010-09-21	11	2075	578	Q4	71	169	83	0.03	0.66	0.87
					Q5	116	219	93	0.05	0.73	0.56
					Q6	186	367	119	0.06	0.80	0.86
					Q7	129	453	114	0.08	0.85	0.82
					…						

1

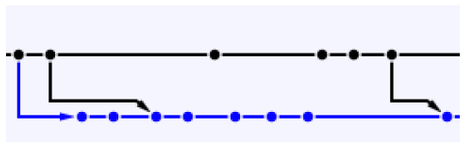
Results

Productivity
(#commits/quarter)

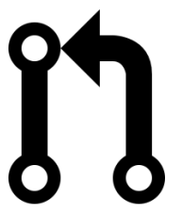


\+ \+ − −

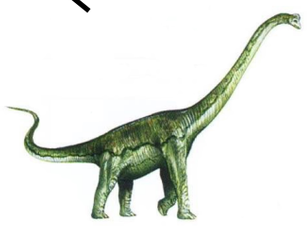
Team size



Overall project activity



Forks



Project age

…

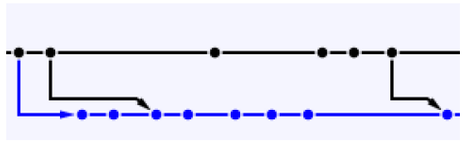
Gender diversity

Tenure diversity

1

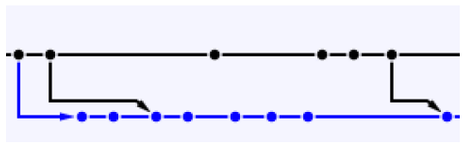
Results

Productivity
(#commits/quarter)

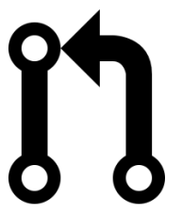


\+ \+ − −

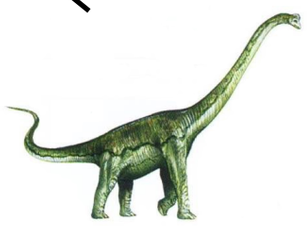
Team size



Overall project activity



Forks



Project age

…

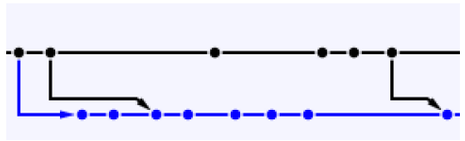
Gender diversity

Tenure diversity

1

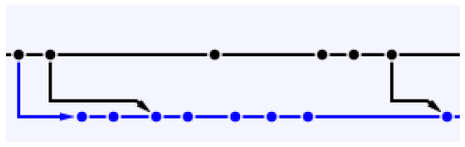
Results

Productivity
(#commits/quarter)

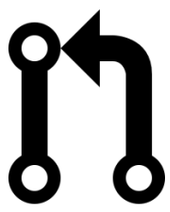


\+ \+ − −

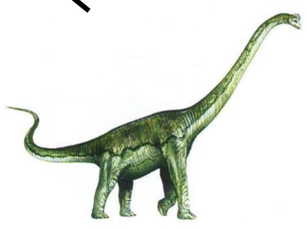
Team size



Overall project activity



Forks



Project age

…

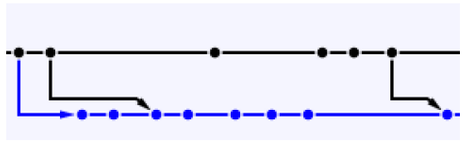
Gender diversity

Tenure diversity

1

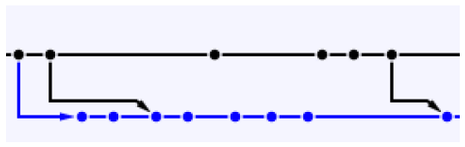
Results

Productivity
(#commits/quarter)

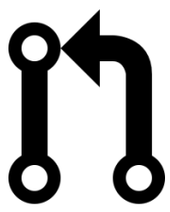


\+ \+ − −

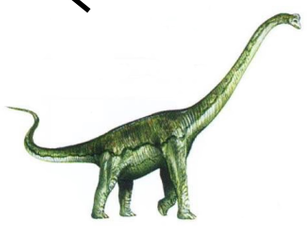
Team size



Overall project activity



Forks



Project age

…

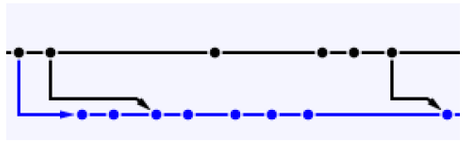
Gender diversity

Tenure diversity

1

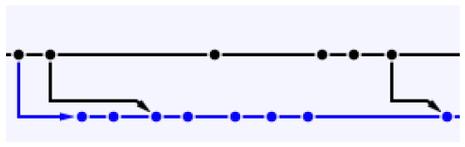
Results

Productivity
(#commits/quarter)

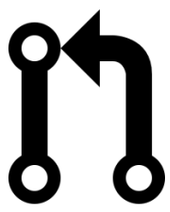


\+ \+ − −

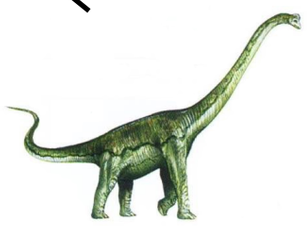
Team size



Overall project activity



Forks



Project age

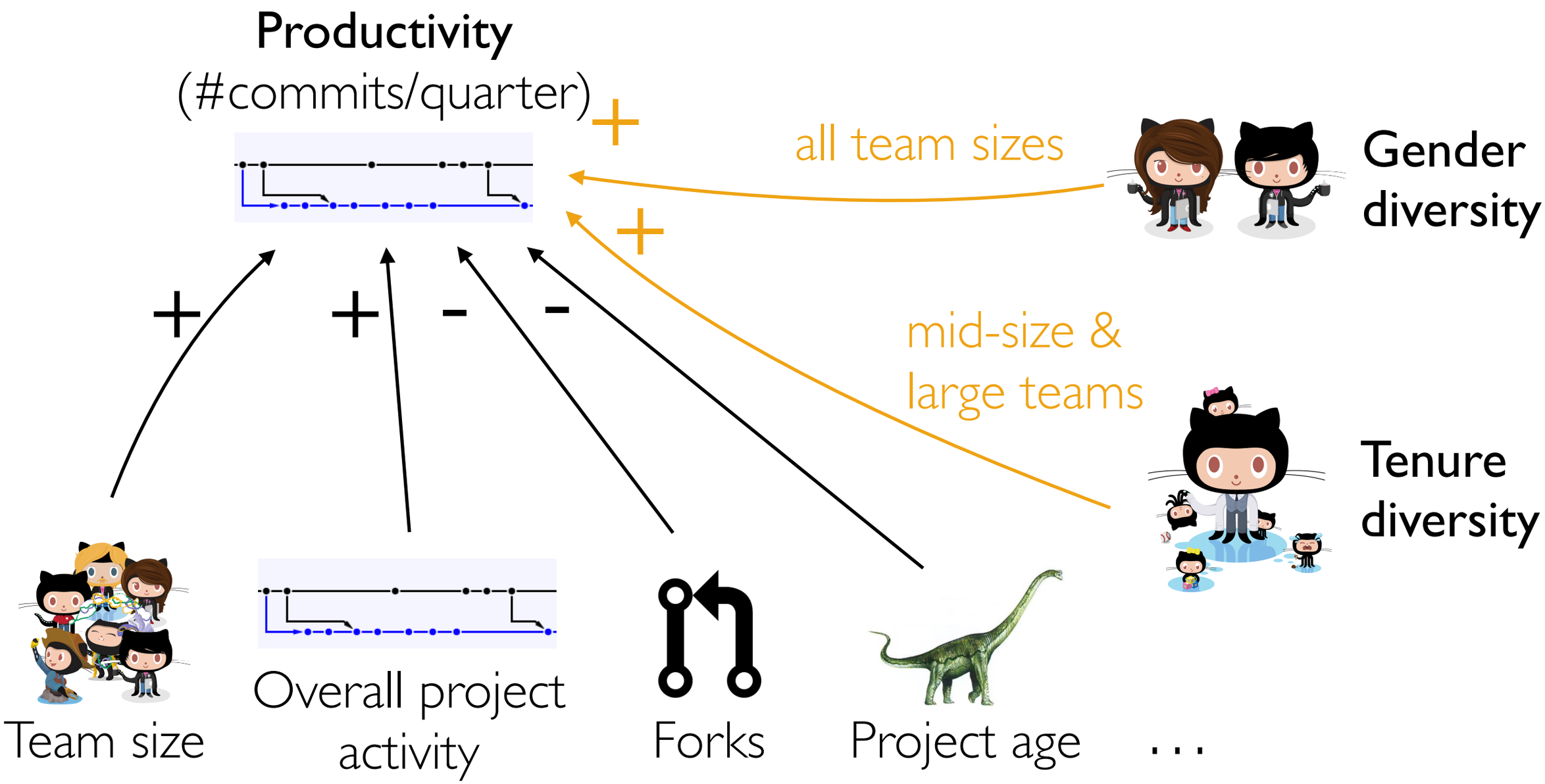
…

Gender diversity

Tenure diversity

1

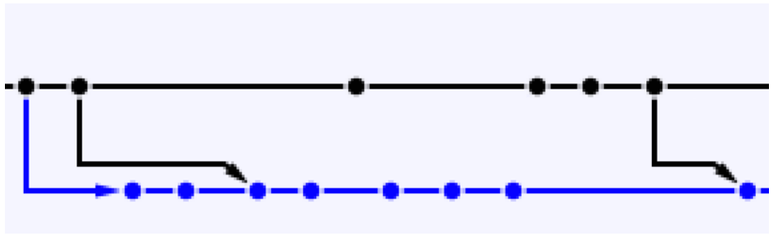
Results



Productivity
(#commits/quarter)
+
+
+
+
-
-
all team sizes
Gender diversity
mid-size & large teams
Tenure diversity
Team size
Overall project activity
Forks
Project age
. . .

Results

Productivity
(#commits/quarter)



+

+

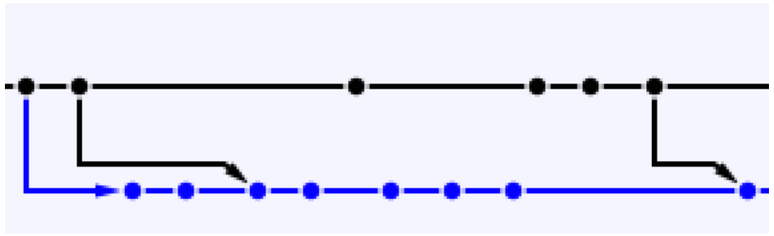
Gender diversity

Tenure diversity

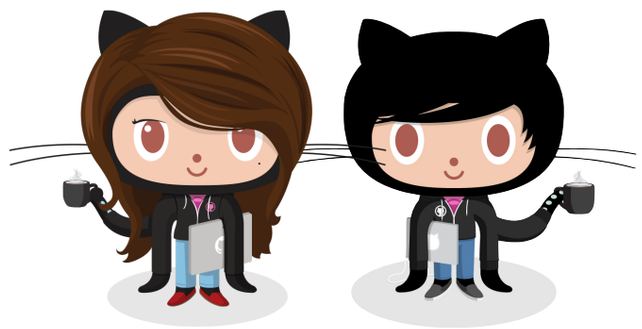
1

Results

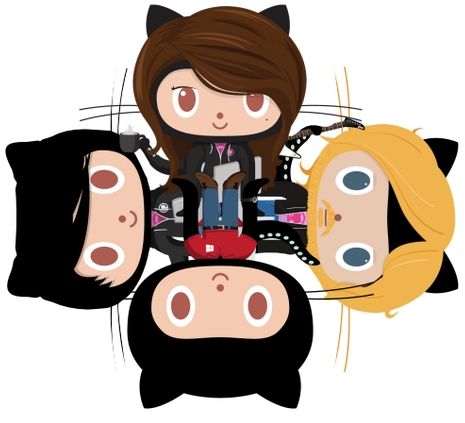
Productivity
(#commits/quarter)



Gender diversity



Turnover (fraction team new w.r.t. prev. quarter)

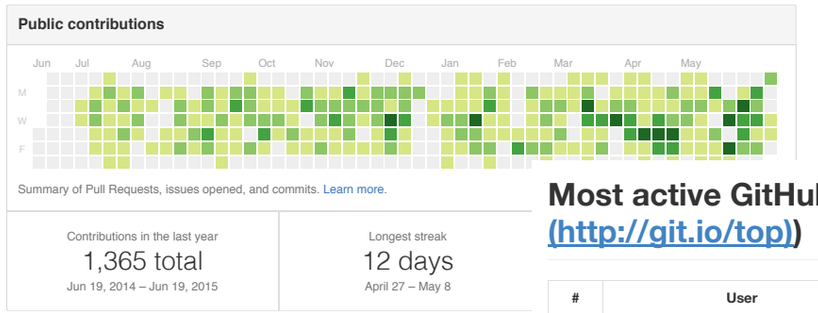


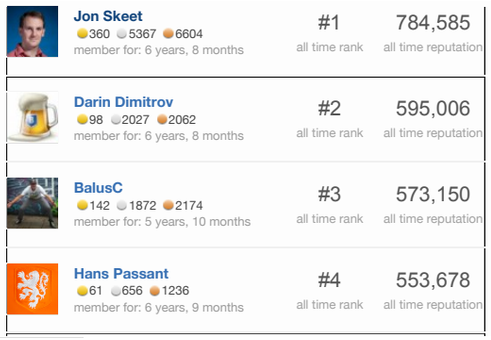
Tenure diversity

1

The takeaway

The social cost of “social coding”





Most active GitHub users (git.io/top (http://git.io/top))

#	User	Contribs	Language	
#1	Ocramius (https://github.com/Ocramius) (Marco Pivetta)	61259	PHP	Frankfurt
#2	michalbe (https://github.com/michalbe) (Michał Budzyński)	14771	JavaScript	Warsaw, Poland
#3	kevinsawicki (https://github.com/kevinsawicki) (Kevin Sawicki)	9159	JavaScript	Redwood City, C

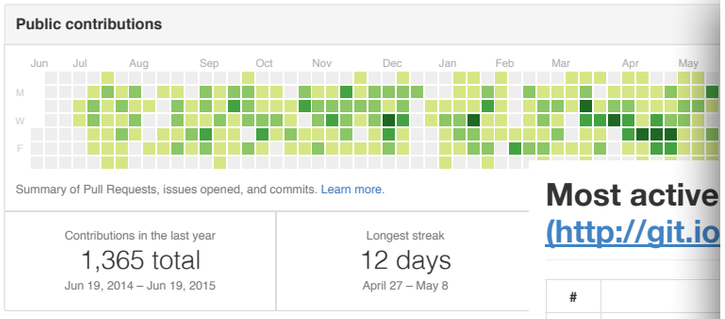
Women are less effective than men in mixed-gender competitive environments [Gneezy et al]

Women shy away from competition and men embrace it [Niederle and Vesterlund]

- Performance in competitive environments: Gender differences
  U Gneezy, M Niederle, A Rustichini. *The Quarterly Journal of Economics, 118(3):1049–1074, 2003*
- Do women shy away from competition? Do men compete too much?
  M Niederle, L Vesterlund. *The Quarterly Journal of Economics, 122(3):1067–1101, 2007*

The takeaway

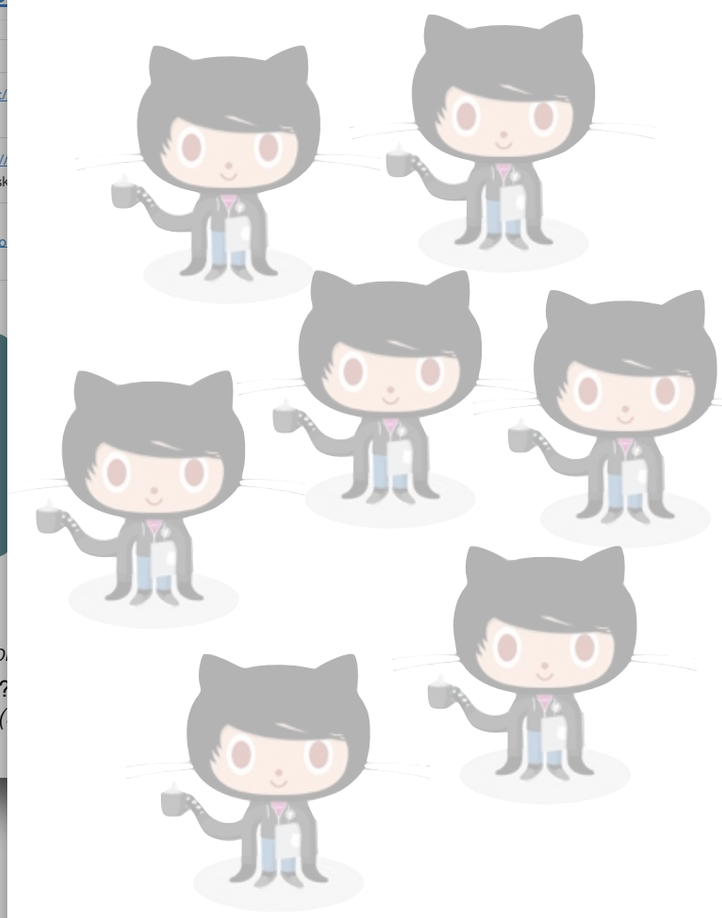
The social cost of “social coding”

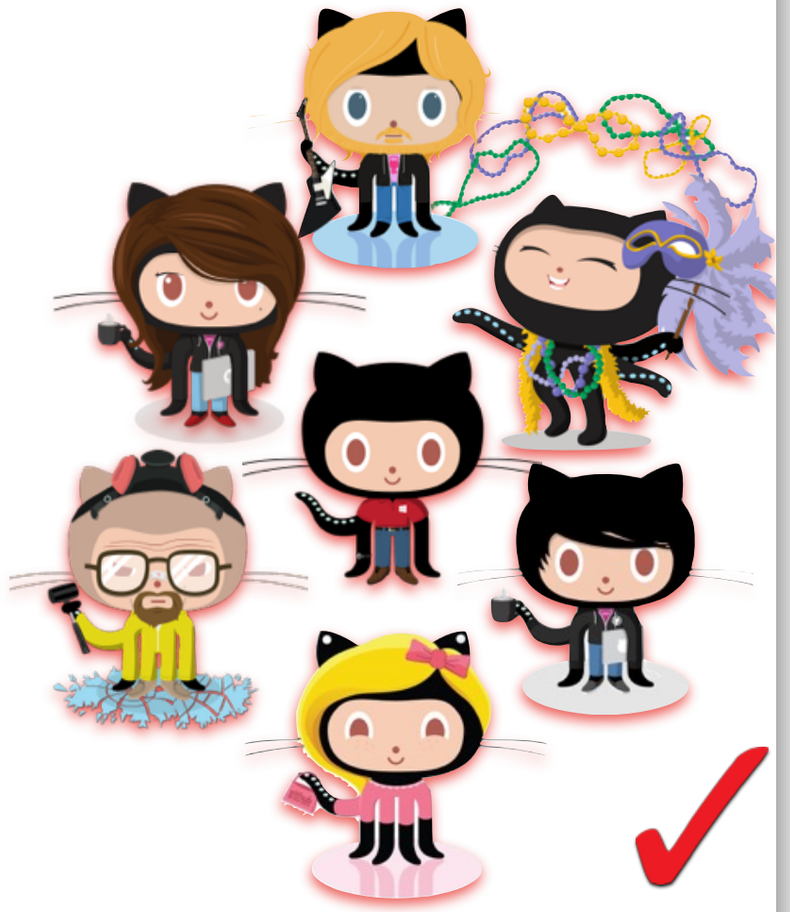


Women are less effective than men in mixed-gender competitive environments [Gneezy et al]

- Performance in competitive environments: Gender differences
  U Gneezy, M Niederle, A Rustichini. *The Quarterly Journal of Econo*
- Do women shy away from competition? Do men compete too much?
  M Niederle, L Vesterlund. *The Quarterly Journal of Economics, 122(*

Which is more effective?





- Gender and tenure diversity in GitHub teams
  B Vasilescu, D Posnett, B Ray, M vd Brand, A Serebrenik, P Devanbu, V Filkov. *CHI 2015*

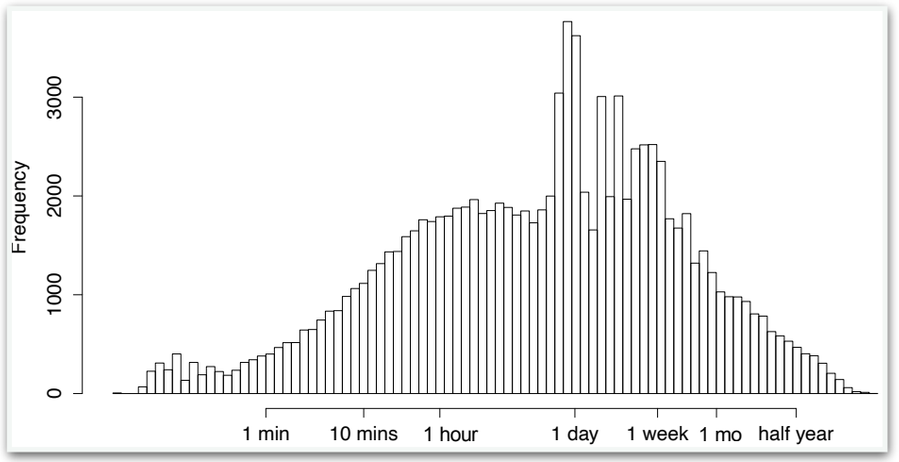
t day

1

2

The social cost of "social coding"

The predictability of "social coding"



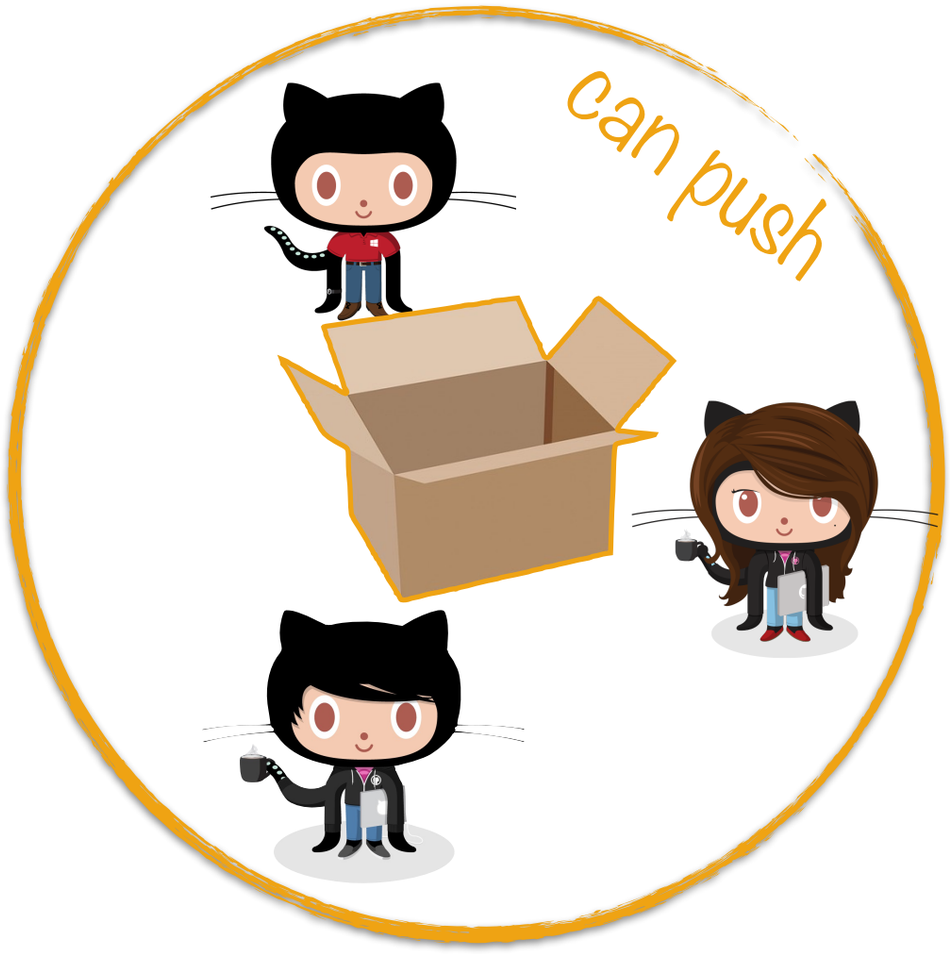
The pull-based model

… traditionally

2

The pull-based model

… traditionally



2

The pull-based model

… traditionally

2

The pull-based model

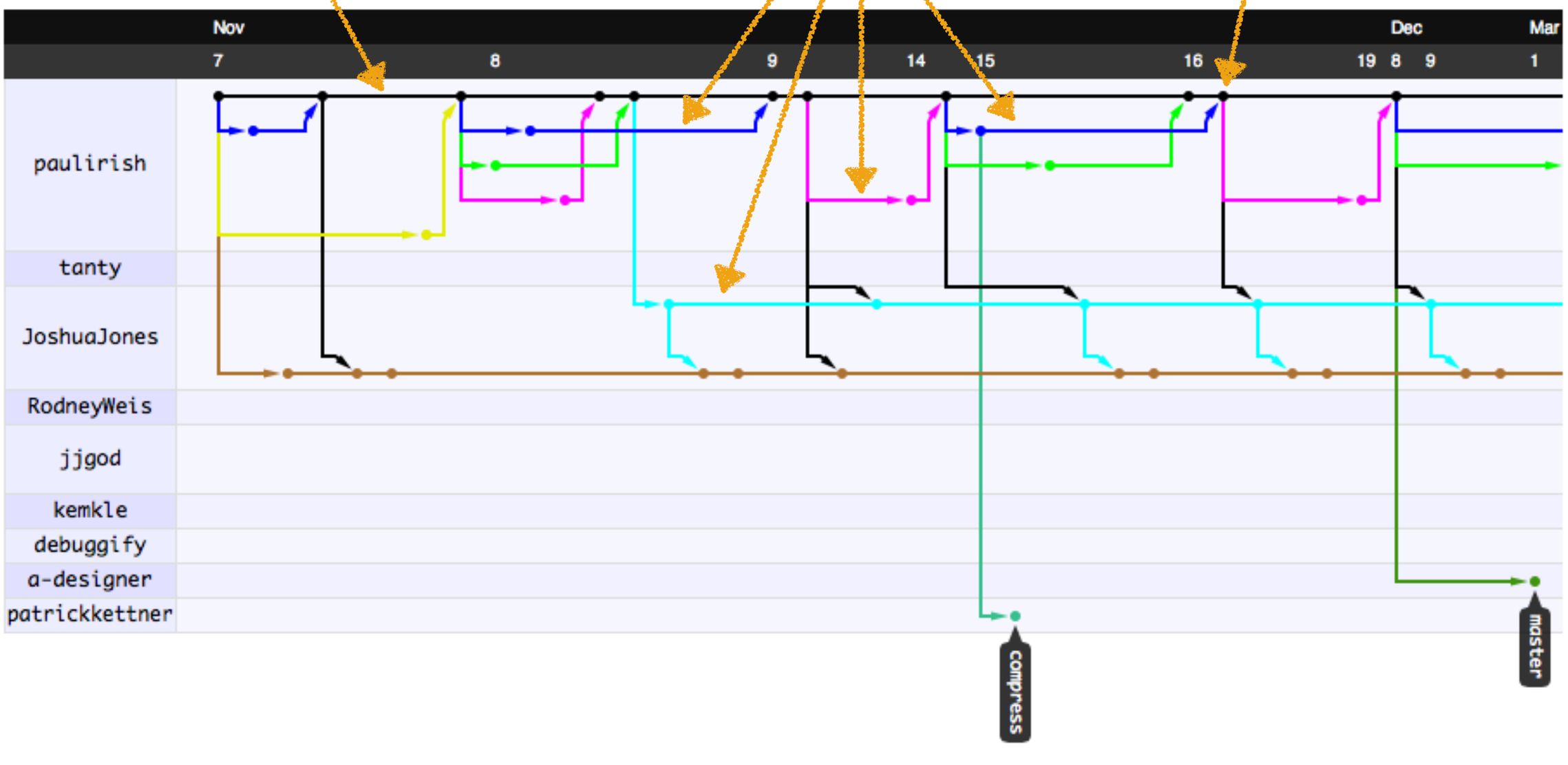
… traditionally

The pull-based model

… traditionally

Main branch

Personal branches / forks



2

The pull-based model

... modernly

The pull-based model

… modernly

- Open source-style collaborative development practices in commercial projects using GitHub
  E Kalliamvakou, D Damian, K Blincoe, L Singer, DM German. *ICSE 2015*
- Work practices and challenges in pull-based development: the integrator's perspective
  G Gousios, A Zaidman, MA Storey, A Van Deursen. *ICSE 2015*

… because code review

Considerable review load

rails / **rails**

Watch 1,887 | Star 26,093 | Fork 10,339

Issues | **Pull requests** | Labels | Milestones

Filters | is:pr is:open | New pull request

467 Open ✔ 12,551 Closed | Author | Labels | Milestones | Assignee | Sort

Move Integer#positive? and Integer#negative? query methods to Numeric ✔ — 2
#20143 opened an hour ago by meinac

Deprecate `assert_template`. ✔ — 8
#20138 opened 9 hours ago by tgxworld

Add Enumerable#map_with to ActiveSupport ✔ — 0
#20134 opened 13 hours ago by mlarraz

Allow creating a save callback for same name with parent association ✔ — 2
#20127 opened 23 hours ago by meinac

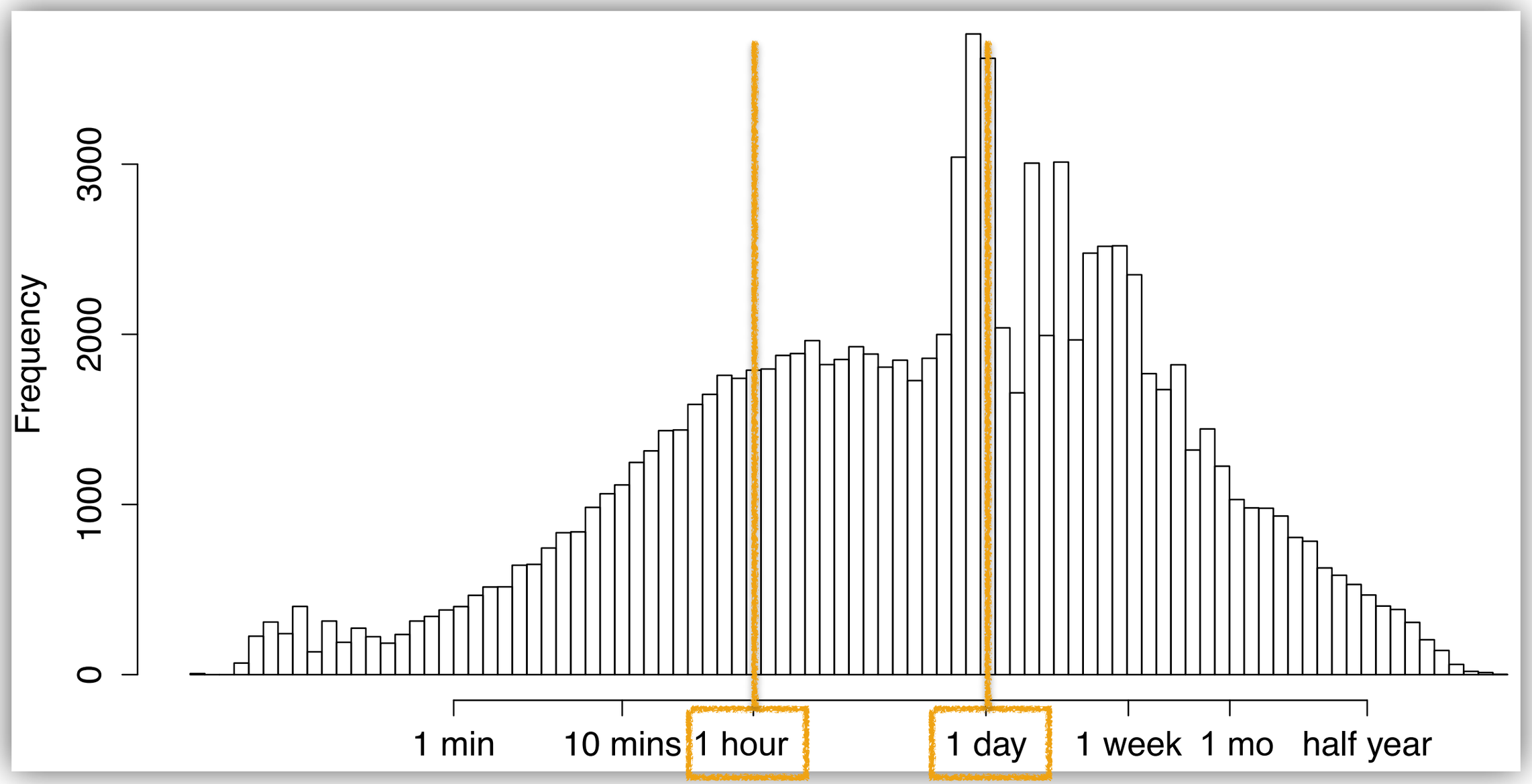
ActiveSupport::HashWithIndifferentAccess select and reject should return enumerator if called without block ✔ — 0
#20125 opened a day ago by imanel

Don't ignore false values for `include_blank` passed to `Tags::Base#select_content_tag` ✔ — 9
#20124 opened a day ago by greysteil

Fix for irregular inflection inconsistency ✔ — 0
#20123 opened a day ago by yoongkang

Add openssl_verify_mode and sync other smtp_settings with API docs ✔ — 0
#20117 opened 2 days ago by jfine

Considerable review load



- Wait for it: Determinants of pull request evaluation latency on GitHub
  Y Yu, H Wang, V Filkov, P Devanbu, B Vasilescu. *MSR 2015*

Automated testing

rails / **rails**

Watch 1,887 | Star 26,093 | Fork 10,339

Issues | Pull requests | Labels | Milestones

Filters | is:pr is:open | New pull request

467 Open ✔ 12,551 Closed — Author | Labels | Milestones | Assignee | Sort

Move Integer#positive? and Integer#negative? query methods to Numeric ✔ — 2
#20143 opened an hour ago by meinac

Deprecate `assert_template` ✔ — 8
#20138 opened 9 hours ago by tgxworld

Add Enumerable#map_with to ActiveSupport ✔ — 0
#20134 opened 13 hours ago by mlarraz

Allow creating a save callback for same name with parent association ✔ — 2
#20127 opened 23 hours ago by meinac

ActiveSupport::HashWithIndifferentAccess select and reject should return enumerator if called without block ✔ — 0
#20125 opened a day ago by imanel

Don't ignore false values for `include_blank` passed to `Tags::Base#select_content_tag ✔ — 9
#20124 opened a day ago by greysteil

Fix for irregular inflection inconsistency ✔ — 0
#20123 opened a day ago by yoongkang

Add openssl_verify_mode and sync other smtp_settings with API docs ✔ — 0
#20117 opened 2 days ago by jfine

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

https://github.com/rails/rails/pull/20064

2

Provide provider_job_id to qu adapter. #20064

Title

Conversation 9 | Commits 1 | Files changed 4

PR Size

kddeisz commented 6 days ago

Submitter

Further work to provide provider_job_id for queue adapters.

Description

https://github.com/rails/rails/pull/20064

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

rafaelfranca commented 6 days ago — Owner

Don't forget the CHANGELOG

https://github.com/rails/rails/pull/20064

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

rafaelfranca commented 6 days ago — Owner

Don't forget the CHANGELOG

kddeisz commented 6 days ago

@rafaelfranca - thanks! Just updated it.

https://github.com/rails/rails/pull/20064

2

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

rafaelfranca commented 6 days ago — Owner

Don't forget the CHANGELOG

kddeisz commented 6 days ago

@rafaelfranca - thanks! Just updated it.

Discussion & code review

jvanbaarsen commented on the diff 5 days ago

activejob/test/integration/queuing_test.rb — View full changes

((12 lines not shown))

```
63  64
    65  +  test 'should supply a provider_job_id when available for delayed jobs' do
    66  +    skip unless adapter_is?(:delayed_job, :sidekiq, :que)
```

jvanbaarsen added a note 5 days ago

I think you forgot to add the provider_job_id for `qu` in the enqueed_at method.

• • •

https://github.com/rails/rails/pull/20064

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

• • •

rafaelfranca commented 6 days ago — Owner

Tests seems to be broken.

https://github.com/rails/rails/pull/20064

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

• • •

rafaelfranca commented 6 days ago — Owner

Tests seems to be broken.

kddeisz commented 6 days ago

It's not displaying the travis errors for me, do you have a link? It was broken until about 20 minutes ago - is it still now? I had to add the "unless qu_job.nil?" because of the different backends.

https://github.com/rails/rails/pull/20064

2

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

• • •

rafaelfranca commented 6 days ago — Owner

Tests seems to be broken.

Automated testing & continuous integration

kddeisz commented 6 days ago

It's not displaying the travis errors for me, do you have a link? It was broken until about 20 minutes ago - is it still now? I had to add the "unless qu_job.nil?" because of the different backends.

• • •

kddeisz commented 5 days ago

Thanks **@matthewd** - looks like it's good now.

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

• • •

rafaelfranca commented 6 days ago — Owner

Tests seems to be broken.

kddeisz commented 6 days ago

It's not displaying the travis errors for me, do you have a link? It was broken until about 20 minutes ago - is it still now? I had to add the "unless qu_job.nil?" because of the different backends.

• • •

kddeisz commented 5 days ago

Thanks **@matthewd** - looks like it's good now.

Merged **rafaelfranca** merged 1 commit into `rails:master` from `kddeisz:qu_provider_job_id` 5 days ago

https://github.com/rails/rails/pull/20064

Provide provider_job_id to qu adapter. #20064

Conversation 9 | Commits 1 | Files changed 4

kddeisz commented 6 days ago

Further work to provide provider_job_id for queue adapters.

• • •

rafaelfranca commented 6 days ago — Owner

Tests seems to be broken.

kddeisz commented 6 days ago

It's not displaying the travis errors for me, do you have a link? It was broken until about 20 minutes ago - is it still now? I had to add the "unless qu_job.nil?" because of the different backends.

• • •

kddeisz commented 5 days ago

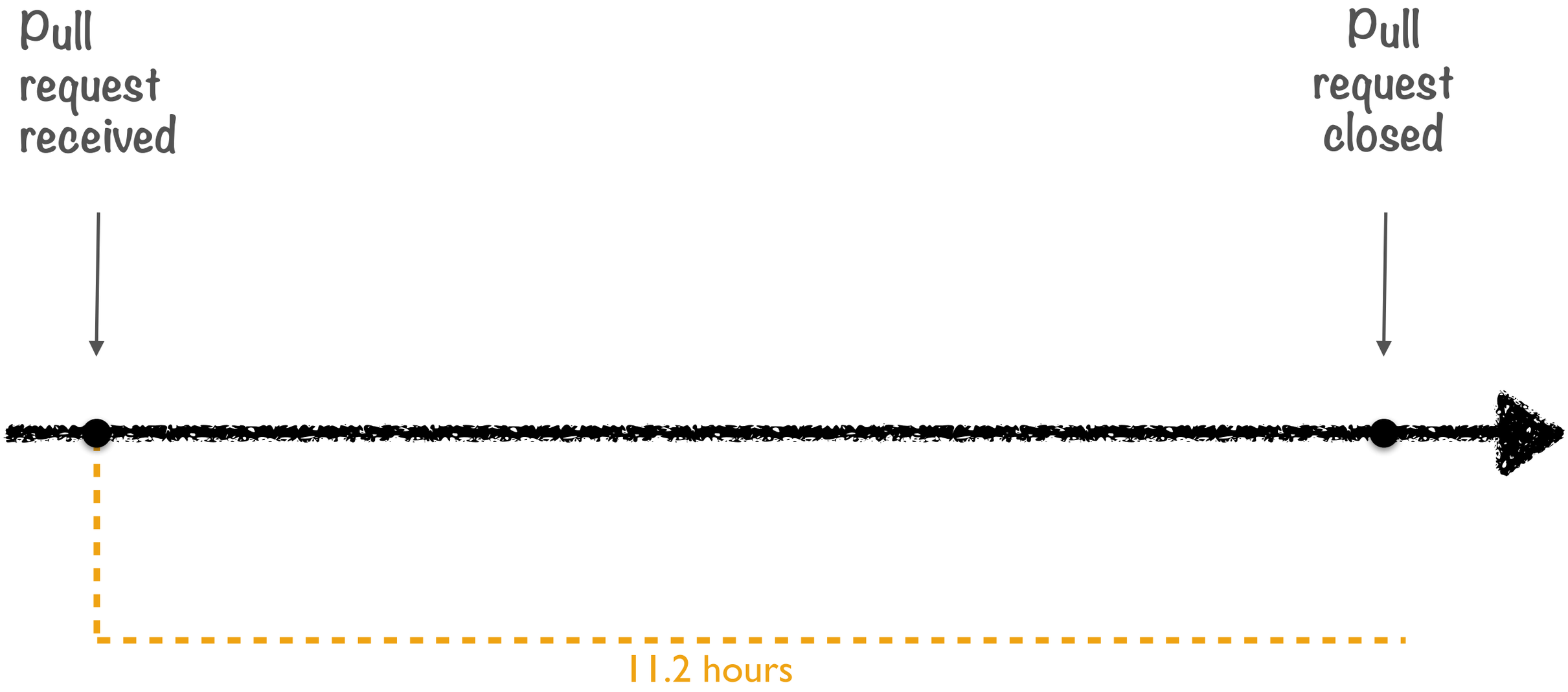
Thanks **@matthewd** - looks like it's good now.

Merged **rafaelfranca** merged 1 commit into `rails:master` from `kddeisz:qu_provider_job_id` 5 days ago

Integrator

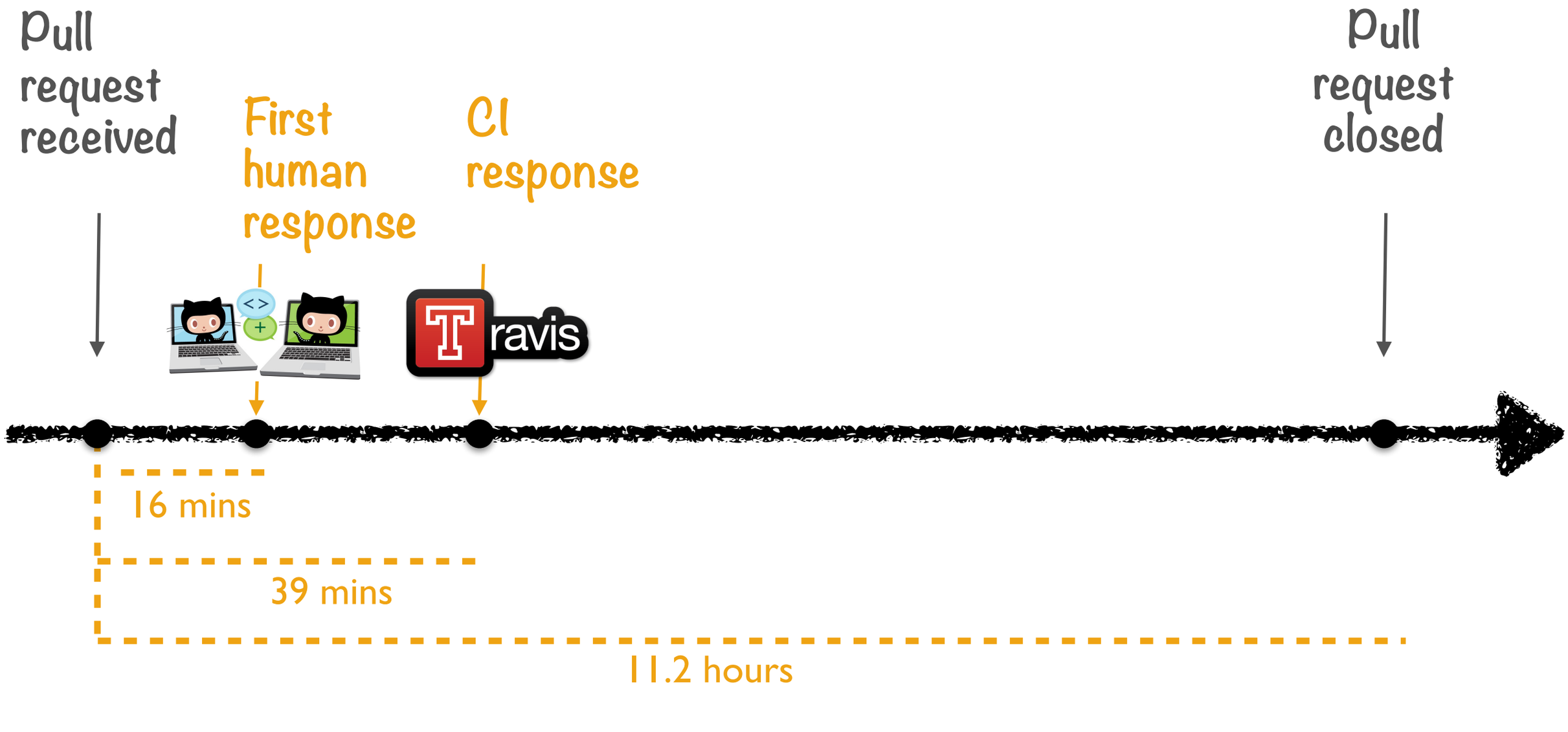
https://github.com/rails/rails/pull/20064

Median pull request times



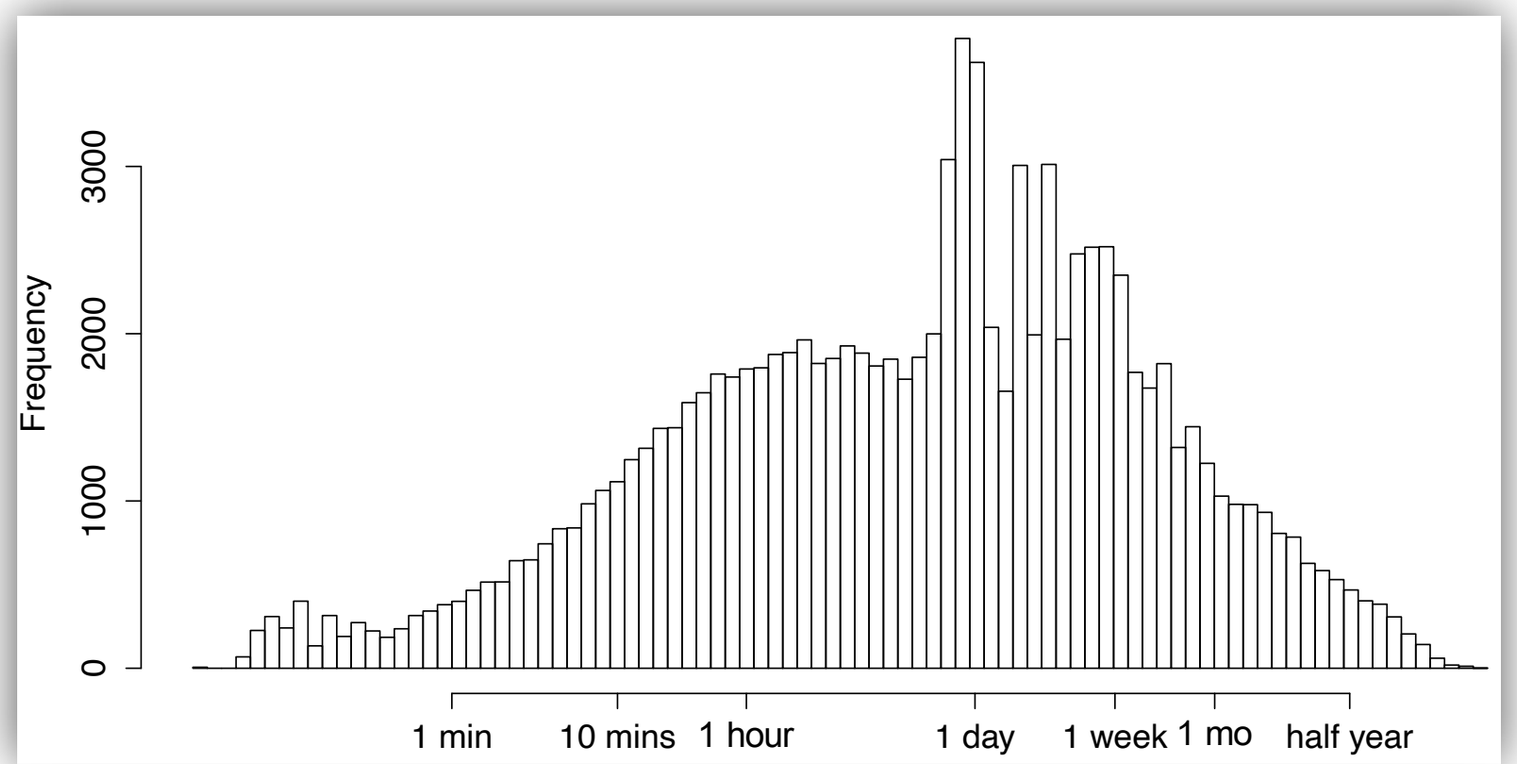
- Wait for it: Determinants of pull request evaluation latency on GitHub
  Y Yu, H Wang, V Filkov, P Devanbu, B Vasilescu. *MSR 2015*

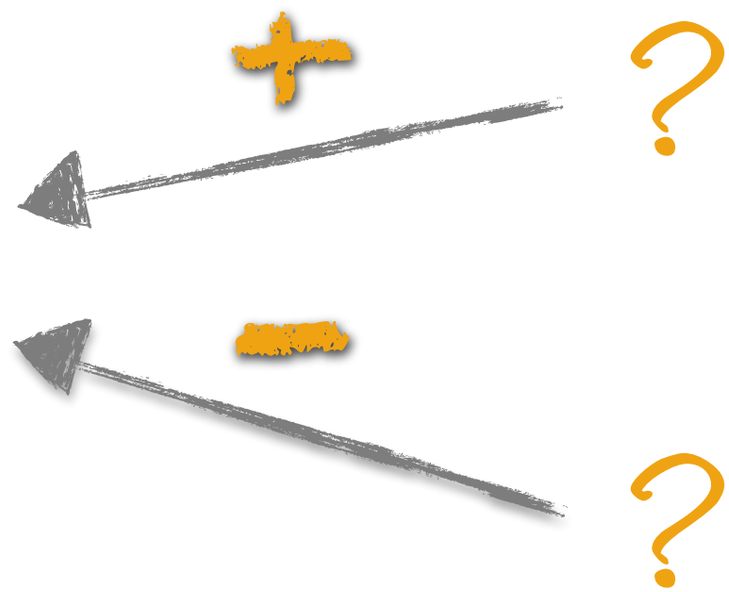
Median pull request times



- Wait for it: Determinants of pull request evaluation latency on GitHub
  Y Yu, H Wang, V Filkov, P Devanbu, B Vasilescu. *MSR 2015*

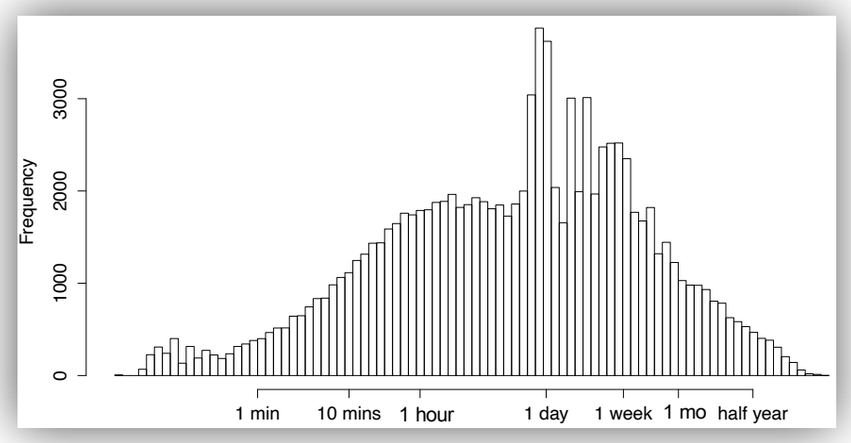
Model PR evaluation times

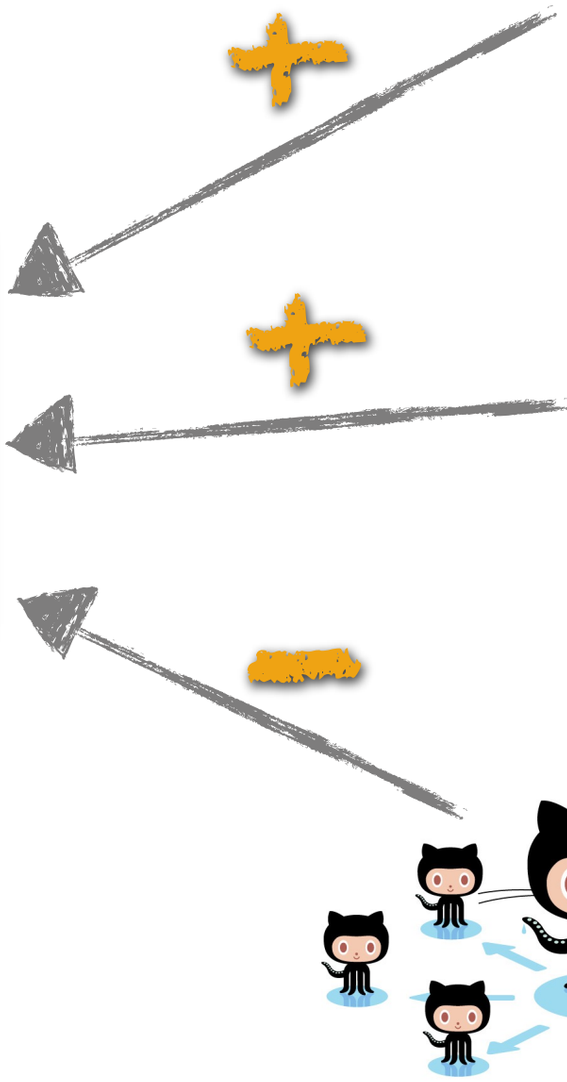


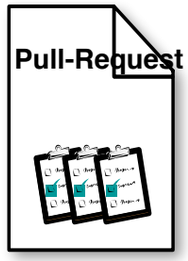


- Wait for it: Determinants of pull request evaluation latency on GitHub
  Y Yu, H Wang, V Filkov, P Devanbu, B Vasilescu. *MSR 2015*

Model PR evaluation times





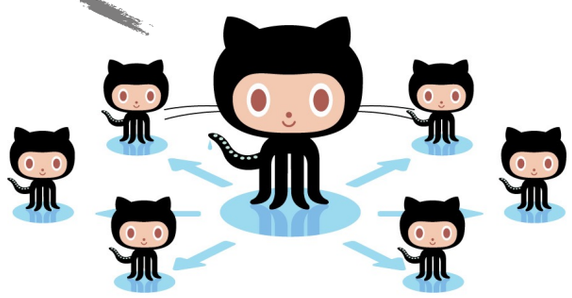


Size

- n_additions
- n_commits

Review

- n_comments



Experience & Social Connections

- merge_rate
- connection_strength
- n_followers

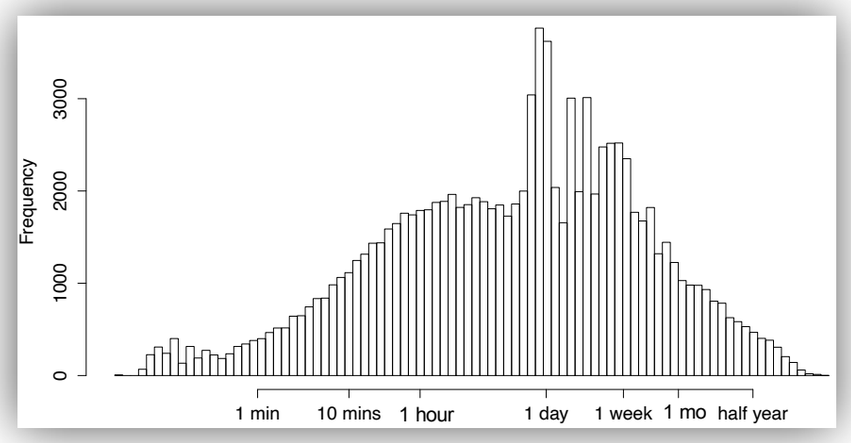
- An exploratory study of the pull-based software development model
  G Gousios, M Pinzger, A Deursen. *ICSE 2014*
- Influence of social and technical factors for evaluating contribution in GitHub
  J Tsay, L Dabbish, J Herbsleb. *ICSE 2014*

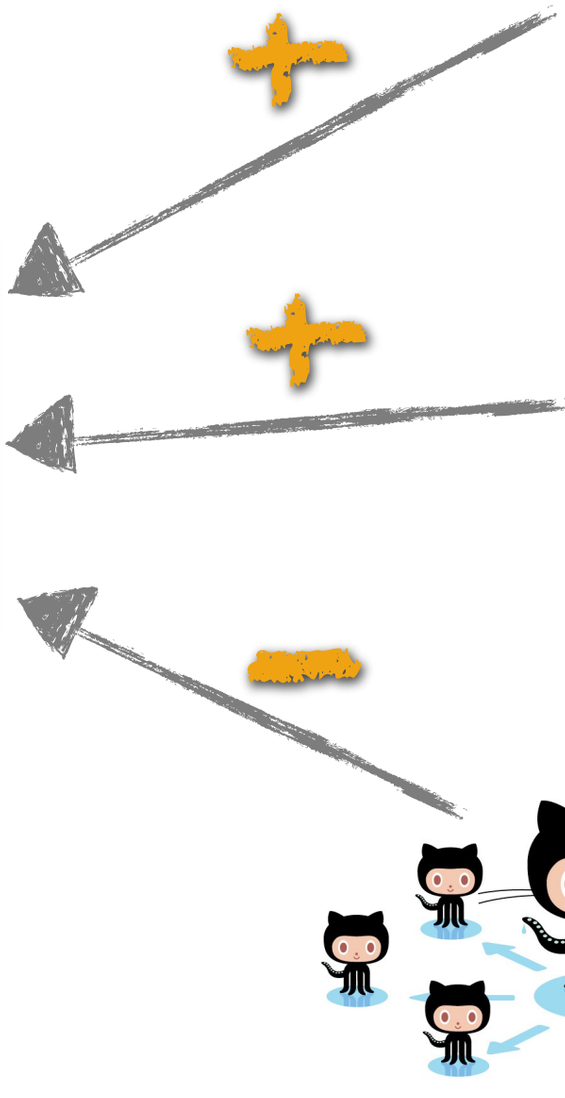
2

Model PR evaluation times

M1: Previously-identified factors

✓ **R^2 = 36.2%**



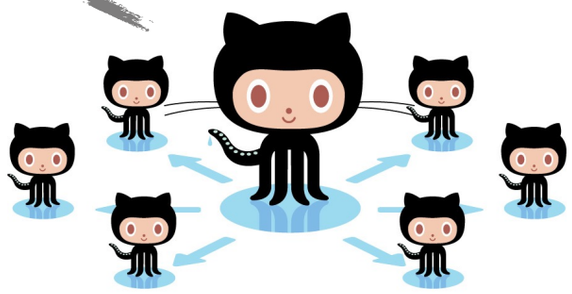


Size

- n_additions
- n_commits

Review

- n_comments



Experience & Social Connections

- merge_rate
- connection_strength
- n_followers

- An exploratory study of the pull-based software development model
  G Gousios, M Pinzger, A Deursen. *ICSE 2014*
- Influence of social and technical factors for evaluating contribution in GitHub
  J Tsay, L Dabbish, J Herbsleb. *ICSE 2014*

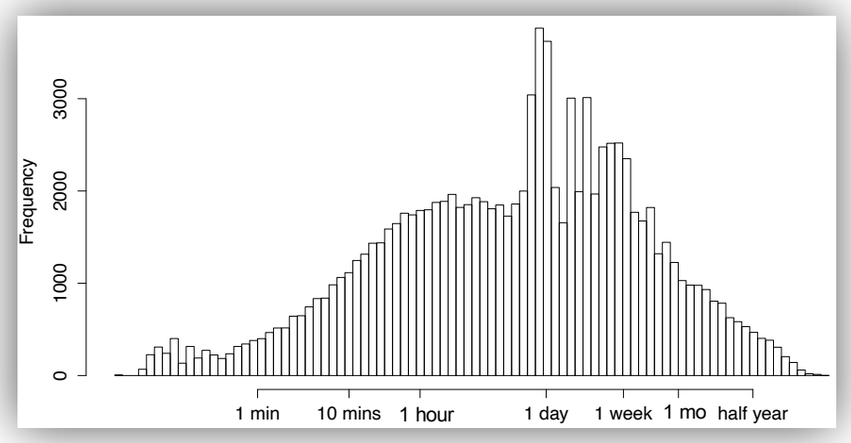
2

Model PR evaluation times

M2: M1 + process-related factors + continuous integration

Title & description

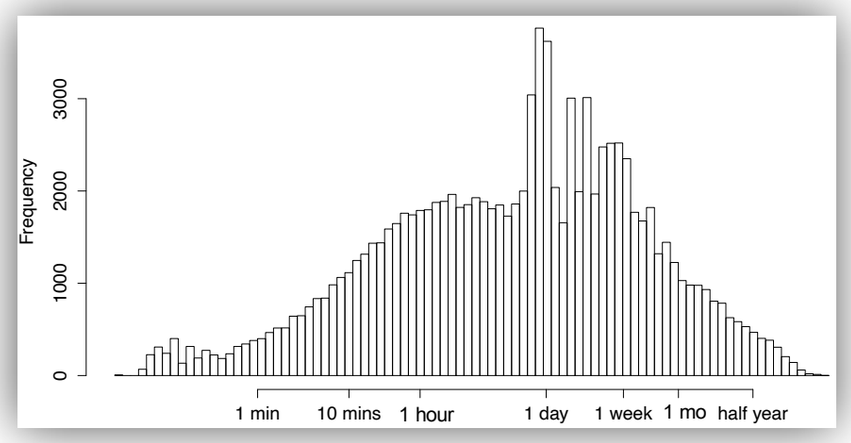
- n_tokens



2

Model PR evaluation times

M2: M1 + process-related factors + continuous integration



Title & description

- n_tokens

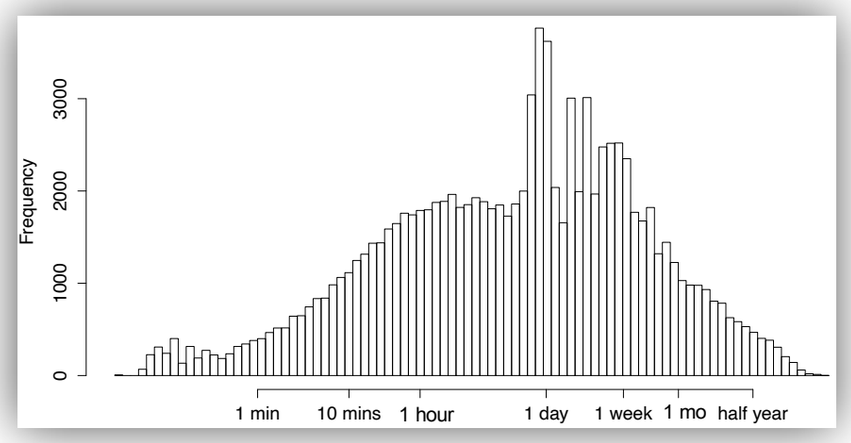
Management

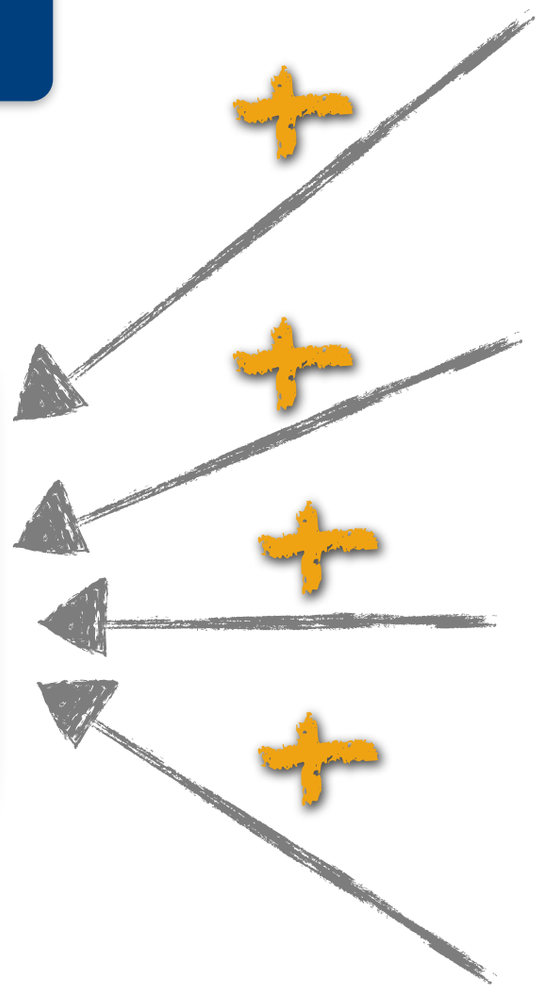
- workload
- availability

2

Model PR evaluation times

M2: M1 + process-related factors + continuous integration





Title & description

- n_tokens

Priority

- time_to_first _response

Continuous Integration

- response time

Management

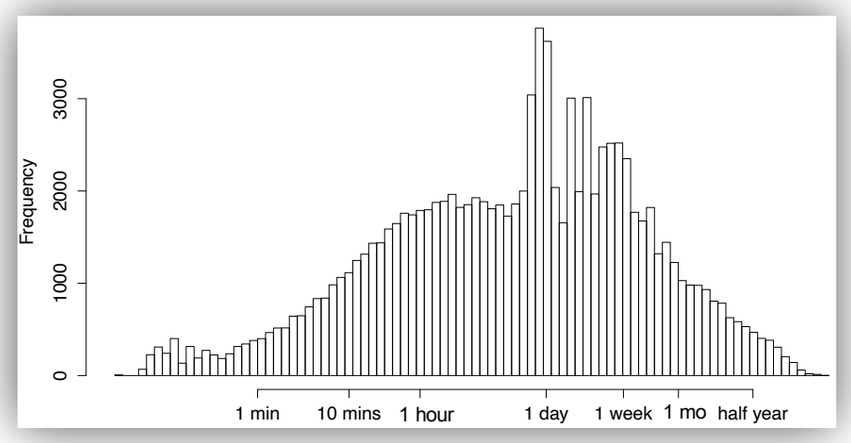
- workload
- availability

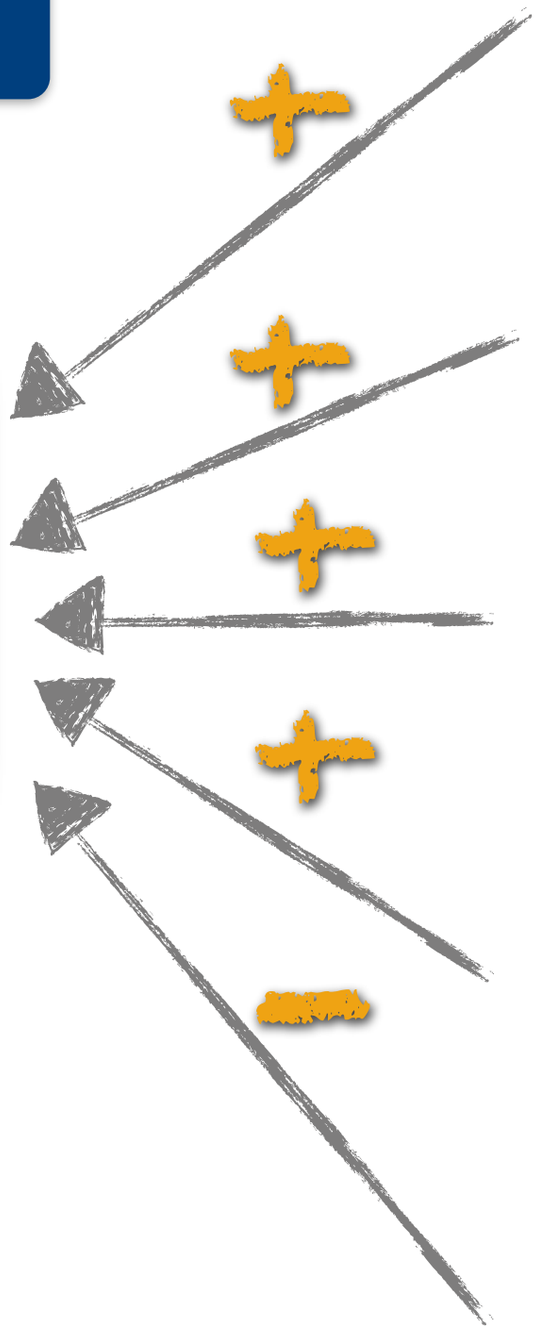
2

Model PR evaluation times

M2: M1 + process-related factors + continuous integration

✓ **R^2 = 58.7%**





Title & description

- n_tokens

Priority

- time_to_first _response

Continuous Integration

- response time

Management

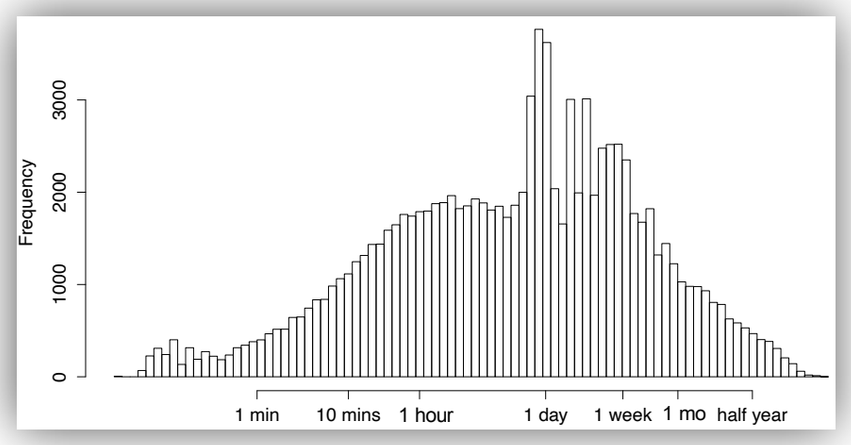
- workload
- availability

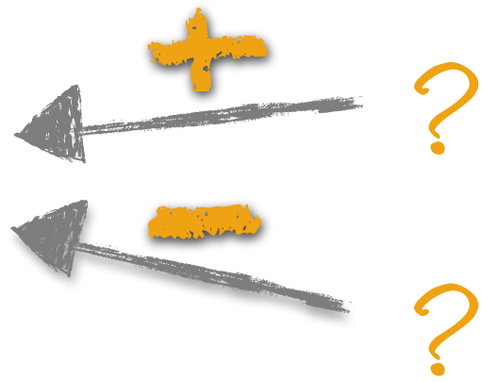
Social tagging

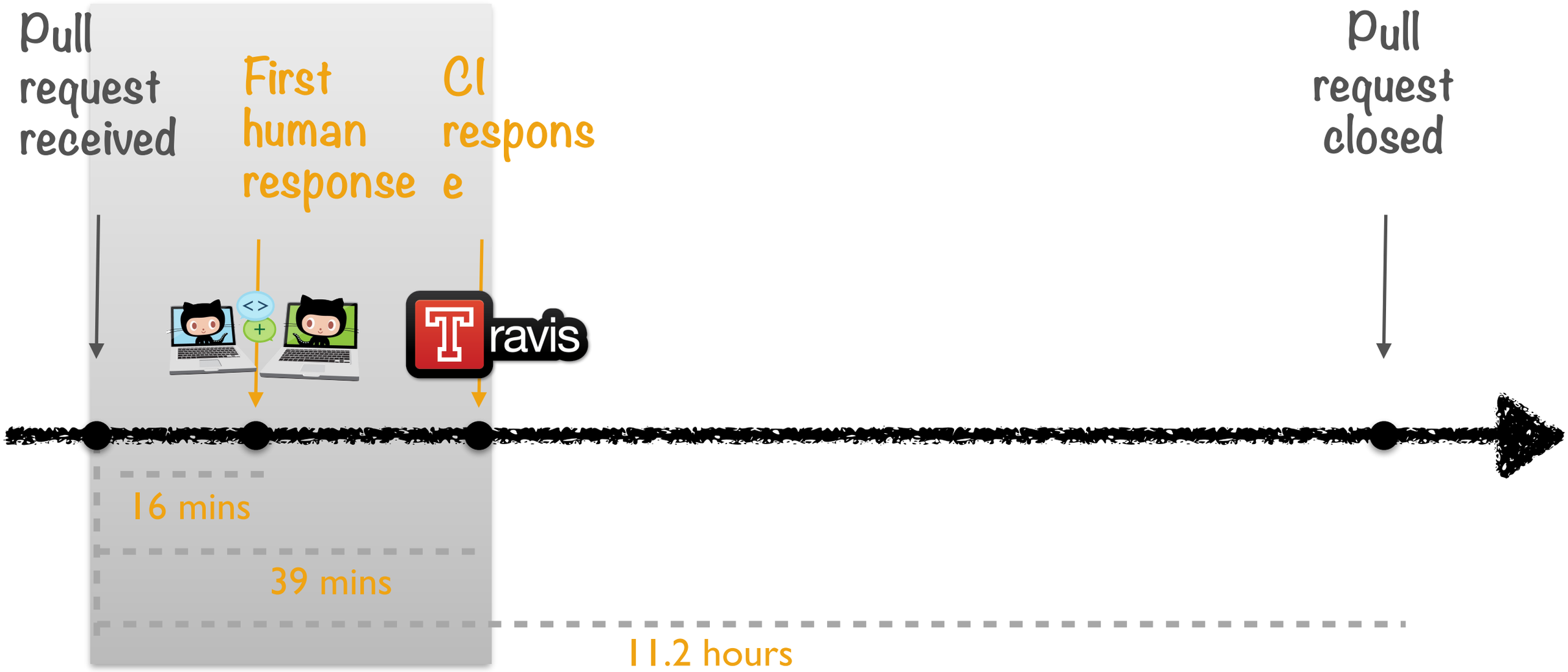
- @mention
- #issue

2

PRs are predictable







Continuous Integration

"[CI] enables us to automate more of our process which frees us up to focus on the important things — like implementing and shipping features! [...]

[The integration of Travis-CI in GitHub] enables the team to rapidly find integration errors or regression failures in the test suite. This tightens the feedback loop and not only enables more defect free code, but greatly speeds up our process."

- Automate All The Things: Continuous Integration and Deployment at RevUnit
  Addam Hardy. http://addamhardy.com/blog/2013/09/28/automate-all-the-things-continuous-integration-and-continuous-deployment-at-revunit/

2

Continuous Integration

Faster process?

Fewer defects?

- Quality and productivity outcomes relating to continuous integration in GitHub
  B Vasilescu,Y Yu, H Wang, P Devanbu, V Filkov. *FSE 2015*

Continuous Integration

Faster process?

+20.5%

more PRs merged monthly from core devs

Fewer defects?

- Quality and productivity outcomes relating to continuous integration in GitHub
  B Vasilescu,Y Yu, H Wang, P Devanbu, V Filkov. *FSE 2015*

2

Continuous Integration

Faster process?

+20.5%

more PRs merged
monthly from core devs

-42.3%

fewer PRs rejected
monthly from core devs

Fewer defects?

-26%

fewer PRs rejected
monthly from externals

- Quality and productivity outcomes relating to continuous integration in GitHub
  B Vasilescu,Y Yu, H Wang, P Devanbu, V Filkov. *FSE 2015*

2

Continuous Integration

Faster process?

+20.5%

more PRs merged
monthly from core devs

-42.3%

fewer PRs rejected
monthly from core devs

-26%

fewer PRs rejected
monthly from externals

Fewer defects?

+48%

more bugs reported
monthly by core devs

-

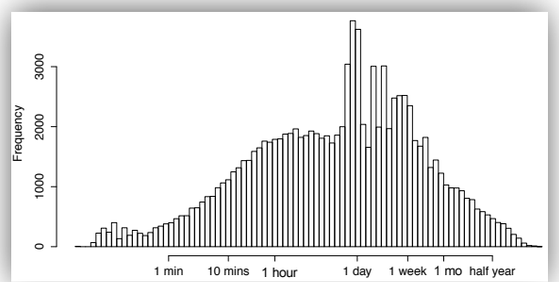
more bugs reported
monthly by externals

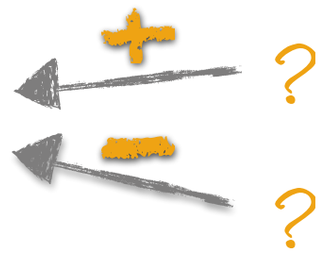
- Quality and productivity outcomes relating to continuous integration in GitHub
  B Vasilescu,Y Yu, H Wang, P Devanbu, V Filkov. *FSE 2015*

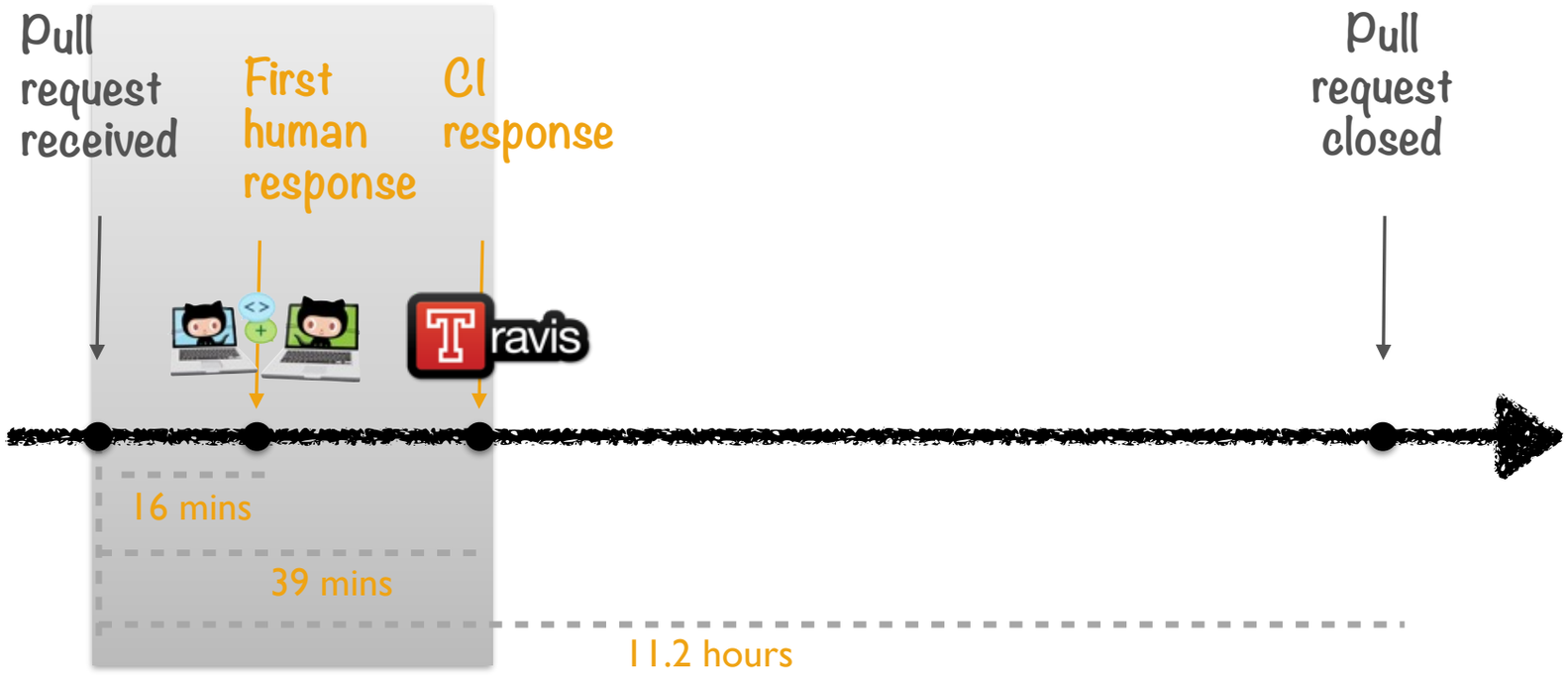
2

The takeaway

PRs are predictable



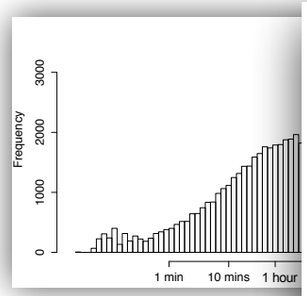


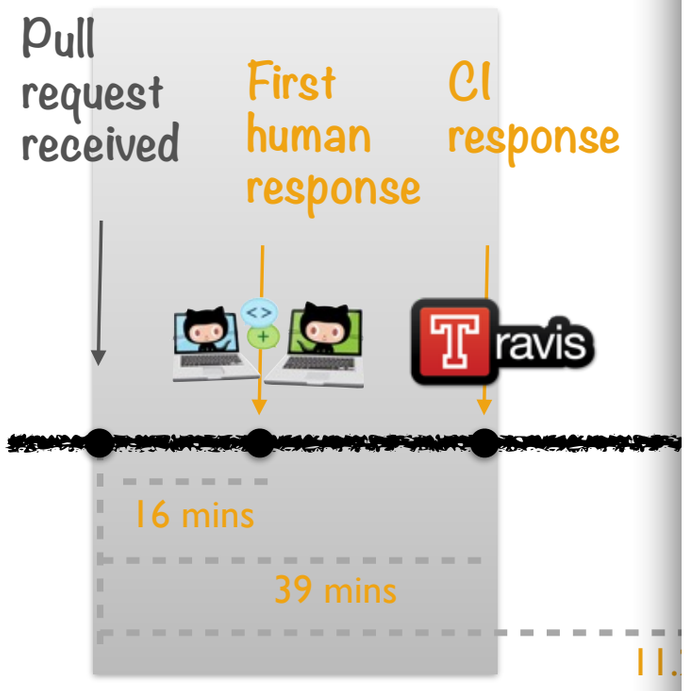


2

The takeaway

PRs are predictable





Continuous Integration

Faster process?

+20.5%

more PRs merged monthly from core devs

-42.3%

fewer PRs rejected monthly from core devs

-26%

fewer PRs rejected monthly from externals

Fewer defects?

+48%

more bugs reported monthly by core devs

-

more bugs reported monthly by externals

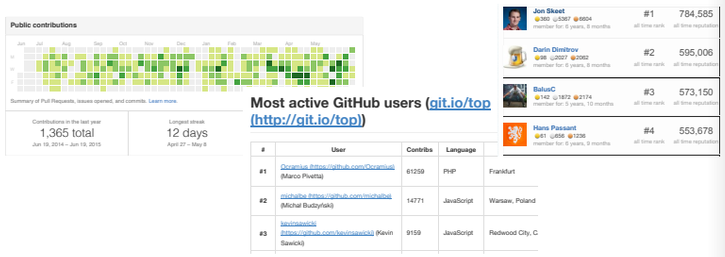
- Quality and productivity outcomes relating to continuous integration in GitHub
  B Vasilescu,Y Yu, H Wang, P Devanbu, V Filkov. *FSE 2015*

OF TEAMS AND AUTOMATION

LESSONS IN SOCIAL CODING FROM GITHUB

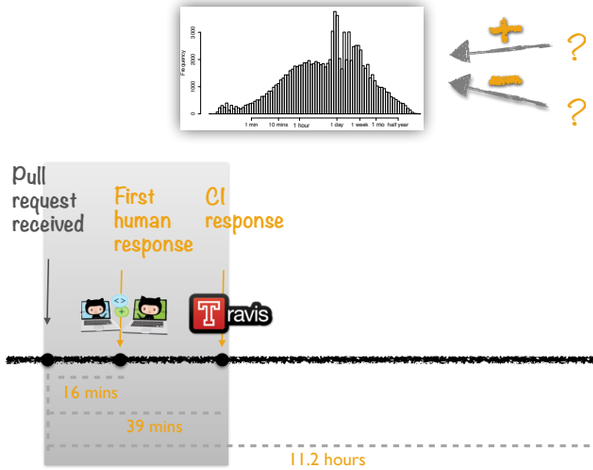
Bogdan Vasilescu

The social cost of “social coding”



- Performance in competitive environments: Gender differences
  U Gneezy, M Niederle, A Rustichini. *The Quarterly Journal of Economics, 118(3):1049–1074, 2003*
- Do women shy away from competition? Do men compete too much?
  M Niederle, L Vesterlund. *The Quarterly Journal of Economics, 122(3):1067–1101, 2007*

PRs are predictable



Which is more effective?

- Gender and tenure diversity in GitHub teams
  B Vasilescu, D Posnett, B Ray, M vd Brand, A Serebrenik, P Devanbu, V Filkov. *CHI 2015*

Continuous Integration

Faster process?

+20.5%
more PRs merged monthly from core devs

-42.3%
fewer PRs rejected monthly from core devs

-26%
fewer PRs rejected monthly from externals

Fewer defects?

+48%
more bugs reported monthly by core devs

-
more bugs reported monthly by externals

- Quality and productivity outcomes relating to continuous integration in GitHub
  B Vasilescu, Y Yu, H Wang, P Devanbu, V Filkov. *FSE 2015*

NWO Netherlands Organisation for Scientific Research
NSF
AFOSR AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
UC DAVIS UNIVERSITY OF CALIFORNIA
TU/e Technische Universiteit Eindhoven University of Technology