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

Hongbo Fang
@fang_hongbo

Hemank Lamba
@hemanklamba

Bogdan Vasilescu
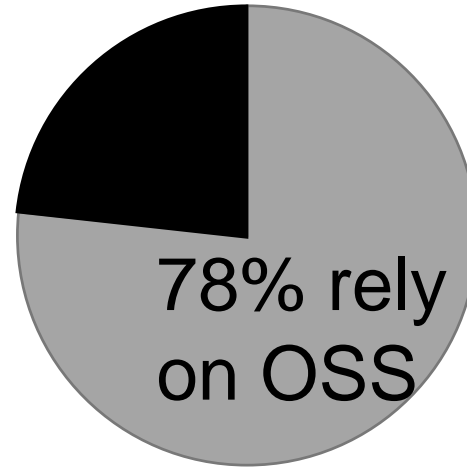@b_vasilescu

James Herbsleb
@jherbsleb

Open source software (OSS) projects are important

Energy

Healthcare

Environment

78% rely on OSS

OSS use by companies

(Black Duck, 2015)

$59B

UK GDP

(Open UK, 2020)

Lack of community attention and contribution are major threats to open source sustainability

Delayed bug fix
(Eghbal et al., 2016)

Project abandonment
(Coelho and Valente, 2017)
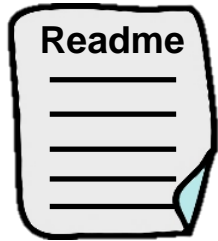
Heartbleed
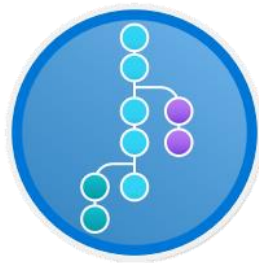(2014)

≠ Important

Fun
(Shah, 2006)

Researchers have studied the effect of several promotion approaches on coding-hosting platforms, but little on others
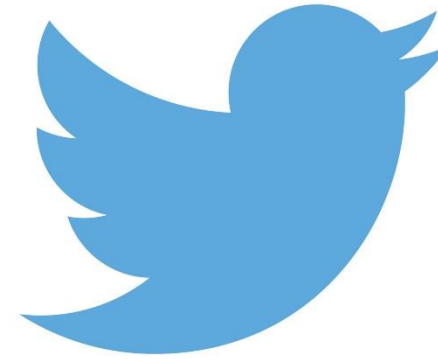
(Trockman et al., 2018)

(Qiu et al., 2019)
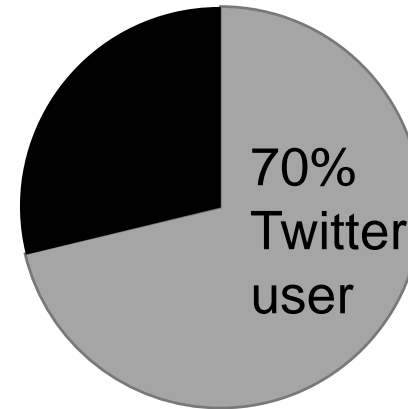
(Borges and Valente, 2018)

?

Twitter is widely used by open source developers for promotion
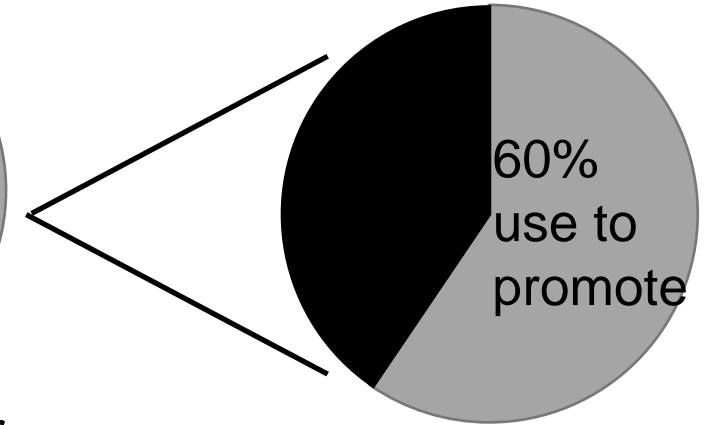
This is a damn slick Node natural language processor -
github.com/axa-group/nlp....

4:14 PM · Mar 14, 2019 · TweetDeck

70%
Twitter
user

OSS developer
Twitter use

60%
use to
promote

Purpose of
Twitter use

(Singer et al., 2014)

Research Questions

1. Whether Twitter promotion on OSS projects help increase project **popularity** (stars) and attract **new developers**?

New developer: Developers who **never** committed to the project **before the tweet**, and made the **first** commit **shortly after**

2. What are factors associated with the size of the impact?

3. What are the characteristics of developers attracted?

Read our paper!! 👉 "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

Result highlight: We claim, causally, that tweet promotion attracts more stars and developers

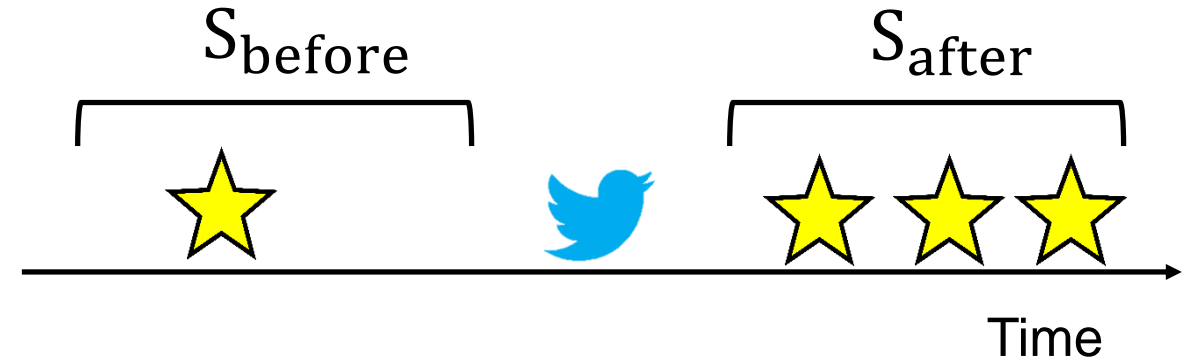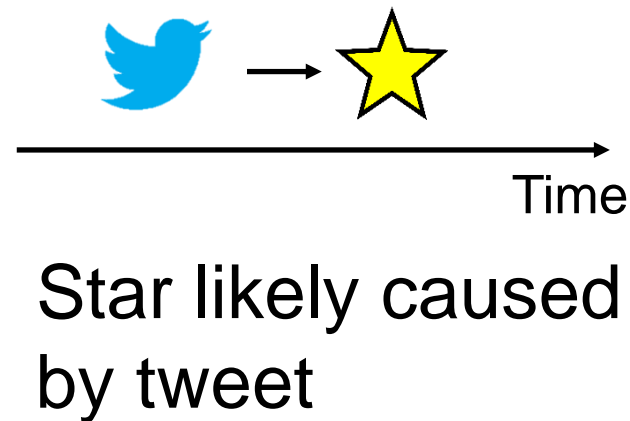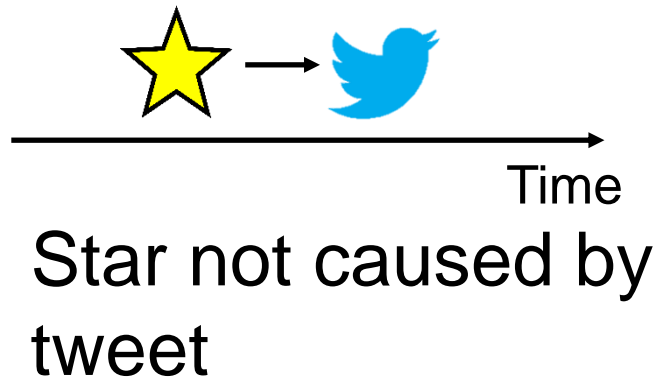Tweet ⟷ Star/developer

Regression doesn't imply causation

Star ~ Tweet
    + Controls
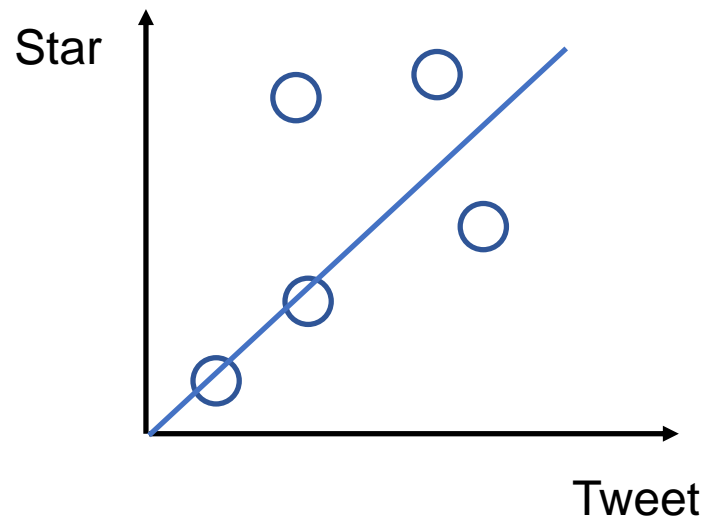
Star caused by tweet = 4 stars ?

Simultaneity bias

Tweet → Star?
Star → Tweet?

Time-series analysis solves simultaneity bias

Star not caused by tweet

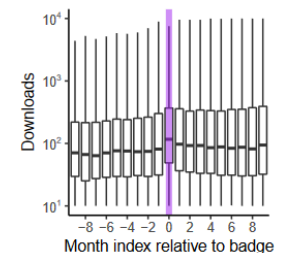Star likely caused by tweet

$$S_{before}$$

$$S_{after}$$

Star caused by tweet

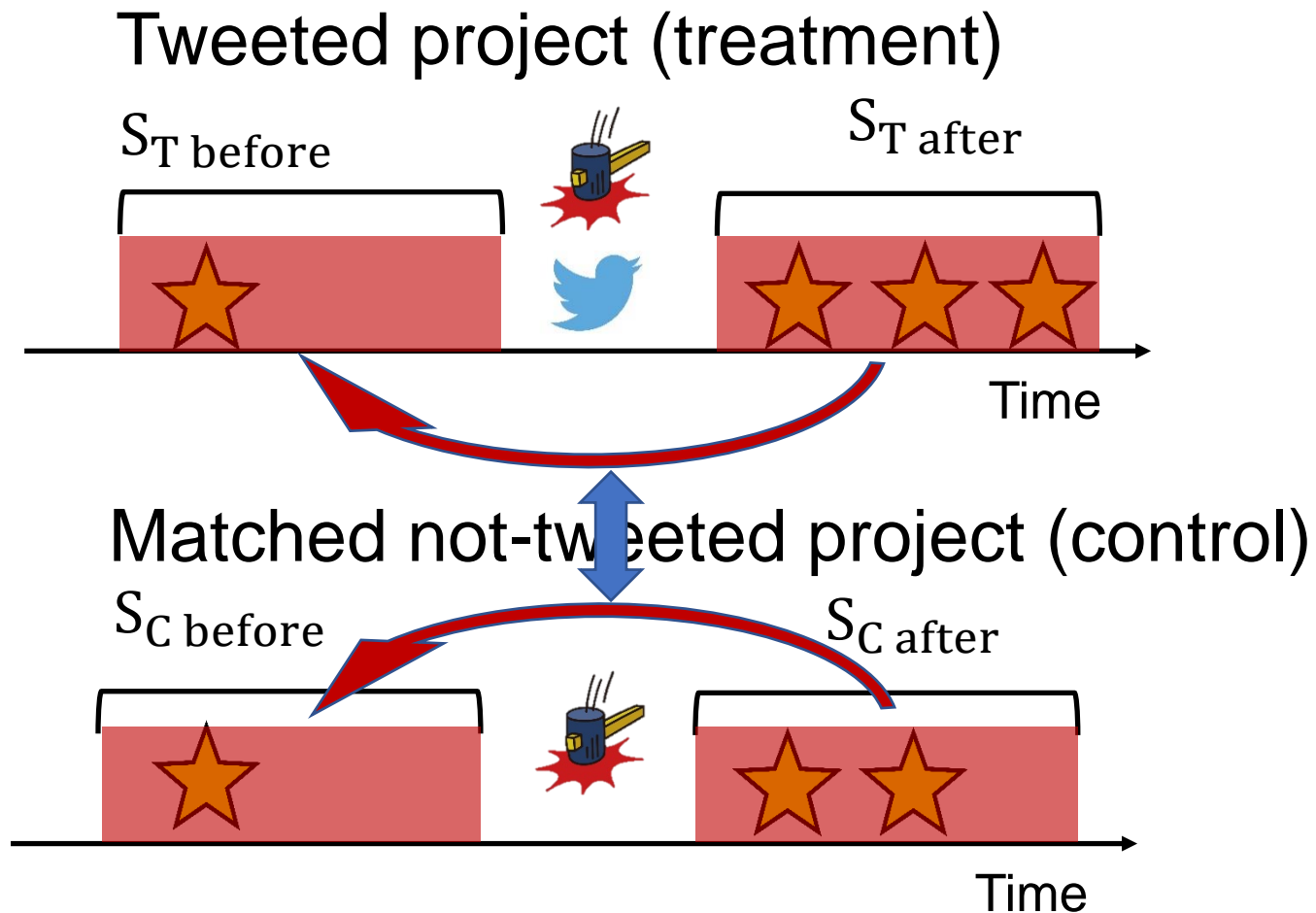$$= S_{after} - S_{before} = 3 - 1 = 2 \ ?$$

Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the *npm* Ecosystem

Asher Trockman,[†] Shurui Zhou,[‡] Christian Kästner,[‡] Bogdan Vasilescu[‡]
[†]University of Evansville, USA   [‡]Carnegie Mellon University, USA

The estimation is not causal because of confounding events, and we use control group to account for their effects

Tweeted project (treatment)

$S_{T \text{ before}}$ $S_{T \text{ after}}$

Time

Matched not-tweeted project (control)

$S_{C \text{ before}}$ $S_{C \text{ after}}$
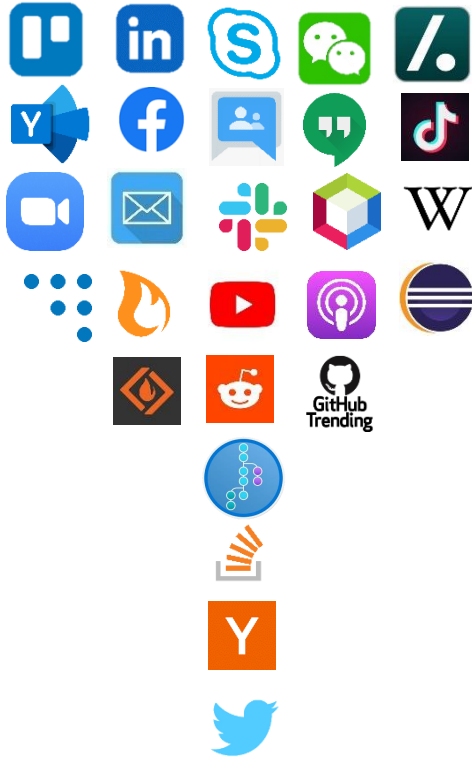
Time

Star caused by tweet

$$= (S_{T \text{ after}} - S_{T \text{ before}}) -$$
$$(S_{C \text{ after}} - S_{C \text{ before}})$$
$$= (3 - 1) - (2 - 1) = 1 \quad ?$$

Difference in Differences

- New feature on GitHub
- COVID
- Financial crisis

We add control variables for confounding events more likely apply to treatment group ($S_{other\ events}$)

GitHub release

GitHub trending

Events on treatment group

Google search result

Our model combines time-series analysis, control group, and a explicit control of confounding events

Star caused by tweet
$\approx (S_{T\ after} - S_{T\ before}) -$
$(S_{C\ after} - S_{C\ before}) -$
$S_{other\ events}$

Difference in Differences
(Dif-in-Dif)

20 + years of test

(Card and Krueger, 1993)

Infer causal relationship with natural experiment, 2021

We conduct a large-scale mixed-method research

Merge tweets into burst

Burst

time

2,370 projects

44,544 tweets

Cross-link developer account on Twitter and GitHub (Fang et al, 2019)

Qualitative analysis

Read our paper!!

RQ1: Tweets attract stars and new developers, the effect is strong on star but marginal on new developers

+7% (+1.2 stars every tweet burst)

+2% (+1 new developer every 250 tweet bursts)

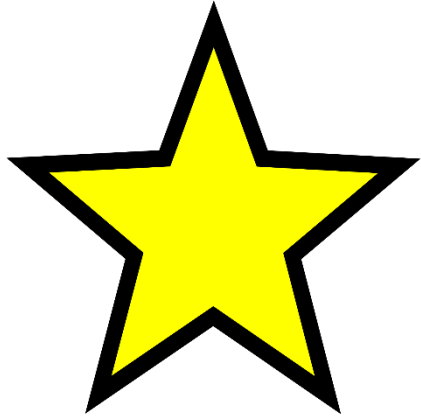# RQ2: Tweet author connection to the project, past Twitter communication, and tweet content matter to the effect size

Sampling CPU and HEAP profiler for #Java featuring AsyncGetCallTrace + perf_events

github.com/jvm-profiling-...

VS

Want to contribute to #tlaplus? Help with 'Sign Toolbox for macOS to remove "unidentified developer" warning'

github.com
Sign Toolbox for macOS to remove the "unidentified develo...
The macOS TLA Toolbox app is not properly signed.
Depending on the macOS' settings the Toolbox either does ...

👉 **Read our paper!!**

Implication 1: Tweet promotion can be a double-edged sword

Draw community attention

Overwhelm developers

Implication 2: Factors associated with the tweet impact provide guidance on tweet promotion

Suggestions

OSS developers

Implication 3: Maintaining social connections and communities on Twitter help attract new developers

OSS projects to be salvaged

Developers connected by Twitter

Implication 4: We have richer data, stronger methods to go beyond the study on code hosting platforms

Richer data

Stronger methods

Our model combines time-series analysis, control group, and a explicit control of confounding events

Star caused by tweet
$\approx (S_{T\ after} - S_{T\ before}) - (S_{C\ after} - S_{C\ before}) - S_{other\ events}$

Difference in Differences (Dif-in-Dif)

20 + years of test
(Card and Krueger, 1993)

NOBEL PRIZE ECONOMICS

Infer causal relationship with natural experiment, 2021

RQ1: Tweets attract stars and new developers, the effect is strong on star but marginal on new developers

+7% (+1.2 stars every tweet burst)

+2% (+1 new developer every 250 tweet bursts)

We conduct a large-scale mixed-method research

Merge tweets into burst
Burst → time

2,370 projects

44,544 tweets

Cross-link developer account on Twitter and GitHub (Fang et al, 2019)

Qualitative analysis — **Read our paper!!**

Implication 1: Tweet promotion can be a double-edged sword
Draw community attention
Overwhelm developers

Implication 2: Factors associated with the tweet impact provide guidance on tweet promotion
Suggestions
OSS developers

Implication 3: Maintaining social connections and communities on Twitter help attract new developers
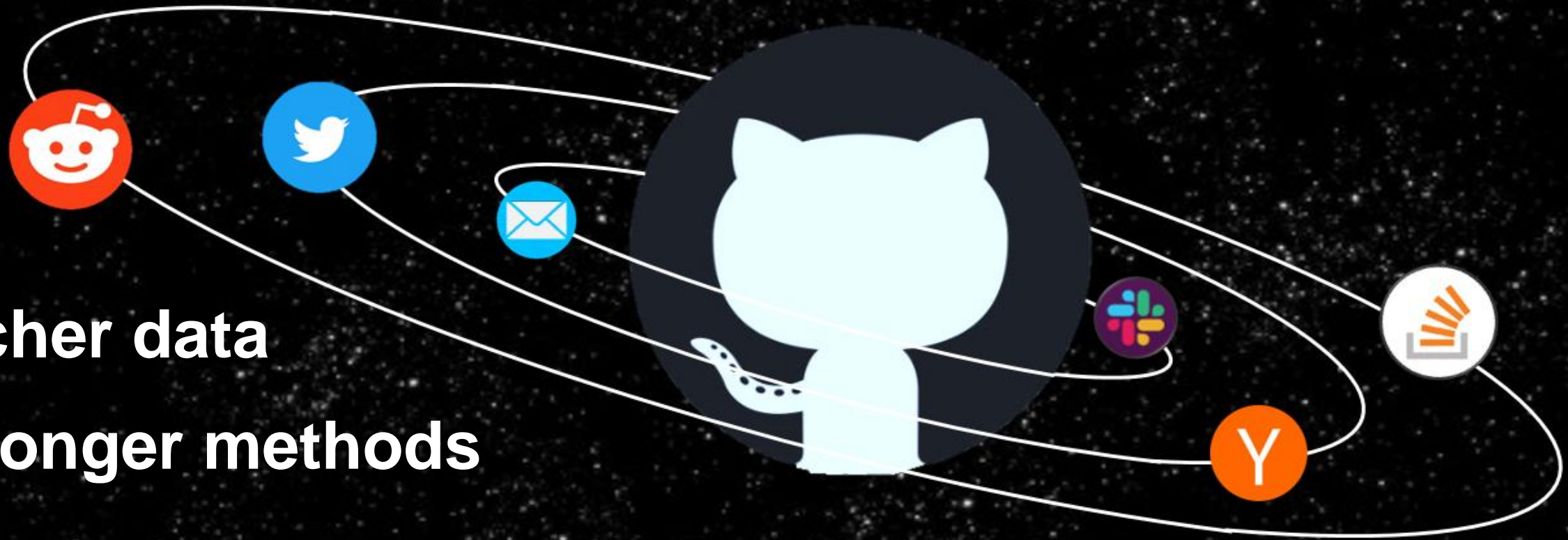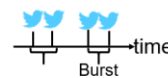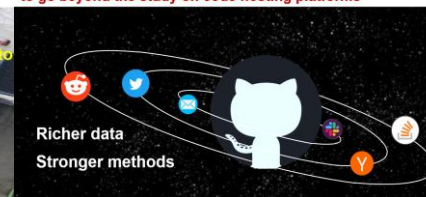Developers connected by Twitter
OSS projects to be salvaged

Implication 4: We have richer data, stronger methods to go beyond the study on code hosting platforms
Richer data
Stronger methods

Paper here!

Replication package!

@fang_hongbo

hongbofa@andrew.cmu.edu