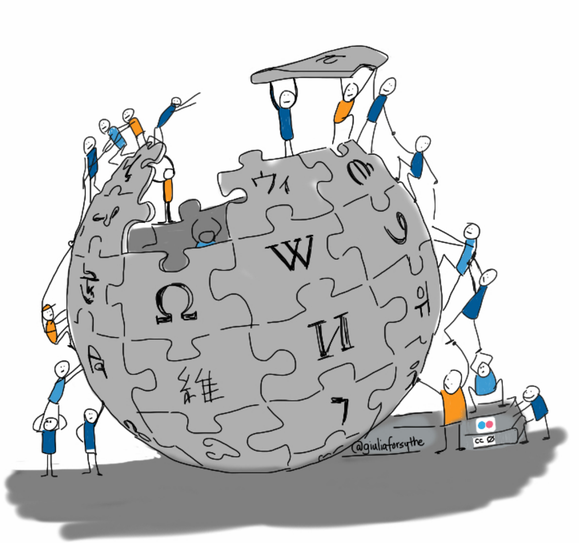
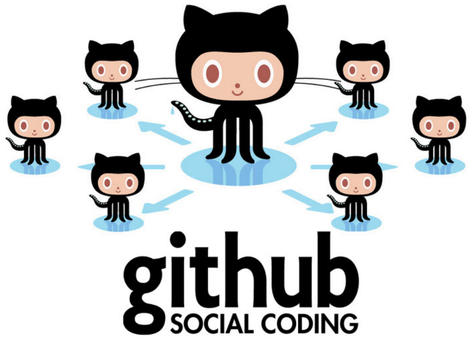
Herding a Deluge of Good Samaritans

How GitHub Projects Respond to Increased Attention

Danaja Maldeniya, Ceren Budak, Lionel P. Robert, Daniel M. Romero
School of Information
University of Michigan

Collaborative crowds produce value in the information economy

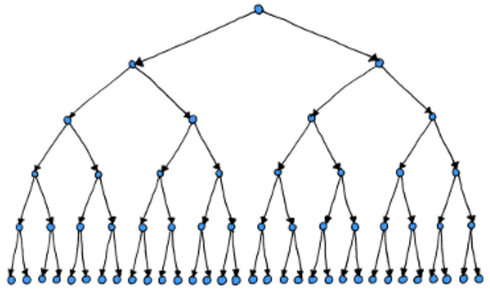




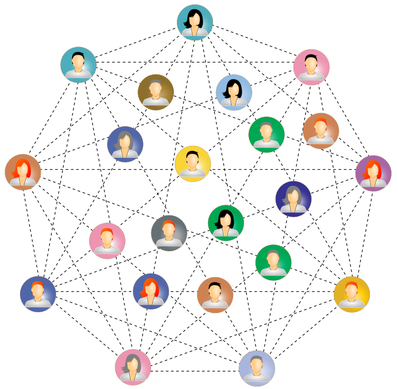
Organizations Vs. Crowds

Organizations Vs. Crowds

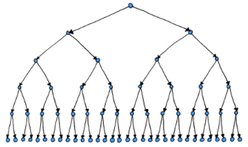
Boundaries and structure



Loosely knit



Organizations Vs. Crowds

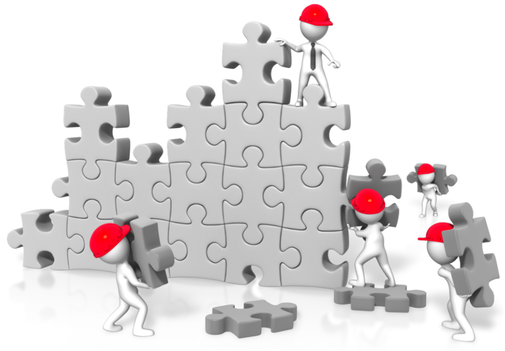


Boundaries and structure

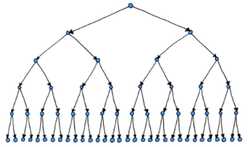
Loosely knit

Formal roles/routines

Flexible roles/routines



Organizations Vs. Crowds



Boundaries and structure

Loosely knit

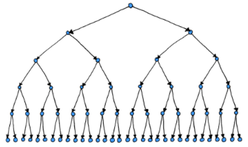
Formal roles/routines

Flexible roles/routines

Contractual

Often non-monetary

Organizations Vs. Crowds



Boundaries and structure

Loosely knit

Formal roles/routines

Flexible roles/routines

Contractual

Often non-monetary

What do these deviations from organizations mean for crowd dynamics?

Research Question

How do crowds respond to exogenous shocks?

Implications

Understanding how crowds respond to shocks and the resulting outcomes can help
(1) teams be better prepared and (2) build better crowd platforms

Current Study

How does a GitHub project trending change the behavior of its core team and the larger contributor community?

Outline

1. Hypotheses
2. GitHub Workflow
3. Shocks on GitHub
4. Causal Framework
5. Findings

Outline

1. **Hypotheses**
2. GitHub Workflow
3. Shocks on GitHub
4. Causal Framework
5. Findings

Hypotheses

Organizational Theory + Empirical Studies of Crowds

Hypotheses

Organizational Theory + Empirical Studies of Crowds

lens: Organizational Change
(Ex: Whetten. 1987)

Hypotheses

Organizational Theory

lens: Organizational Change (Ex: Whetten. 1987)

setting: Rapidly growing small organizations (Fombrun and Wally. 1989; Hambrick and Crozier. 1985 etc.)

+

Empirical Studies of Crowds

Hypotheses

Organizational Theory + Empirical Studies of Crowds

lens: Organizational Change
(Ex: Whetten. 1987)

setting: Rapidly growing small organizations (Fombrun and Wally. 1989; Hambrick and Crozier. 1985 etc.)

rationale:

1. long-tail size and popularity distribution of GitHub projects
2. Member vs. Outsider dichotomy

Hypotheses

Organizational Theory

lens: Organizational Change (Ex: Whetten. 1987)

setting: Rapidly growing small organizations (Fombrun and Wally. 1989; Hambrick and Crozier. 1985 etc.)

rationale:

1. long-tail size and popularity distribution of GitHub projects
2. Member vs. Outsider dichotomy

+

Empirical Studies of Crowds

GitHub Dynamics:
Contribution (Ex: Dabbish et al. 2012)
Coordination (Ex: Romero et al. 2015)
Success (Ex: Vasilescu et al. 2015)
etc.

Hypotheses

Organizational Theory

lens: Organizational Change (Ex: Whetten. 1987)

setting: Rapidly growing small organizations (Fombrun and Wally. 1989; Hambrick and Crozier. 1985 etc.)

rationale:

1. long-tail size and popularity distribution of GitHub projects
2. Member vs. Outsider dichotomy

+

Empirical Studies of Crowds

GitHub Dynamics:
Contribution (Ex: Dabbish et al. 2012)
Coordination (Ex: Romero et al. 2015)
Success (Ex: Vasilescu et al. 2015)
etc.

Shocks on Crowds:
On Wikipedia;
(1) Loss of workforce (Zhang et al. 2017)
(2) Recognition (Zhang et al. 2018)
(3) Attention (Zhang et al, 2019)

Hypotheses

1. **Growth** — Trending increase community engagement (Begel et al. 2013, Jiang et al. 2017)

Hypotheses

1. **Growth** — Trending increase community engagement
   (Begel et al. 2013, Jiang et al. 2017)

2. **Growing Pains** —Strain on coordination will cause backlogs
   (Fombrun and Wally. 1989, Hambrick and Crozier. 1985)

Hypotheses

1. **Growth** — Trending increase community engagement (Begel et al. 2013, Jiang et al. 2017)

2. **Growing Pains** —Strain on coordination will cause backlogs (Fombrun and Wally. 1989, Hambrick and Crozier. 1985)

3. **Work Routines** — Core team will transition to admin. role (Child and Keiser. 1981, Hambrick and Crozier. 1985)

Hypotheses

1. **Growth** — Trending increase community engagement
   (Begel et al. 2013, Jiang et al. 2017)

2. **Growing Pains** —Strain on coordination will cause backlogs
   (Fombrun and Wally. 1989, Hambrick and Crozier. 1985)

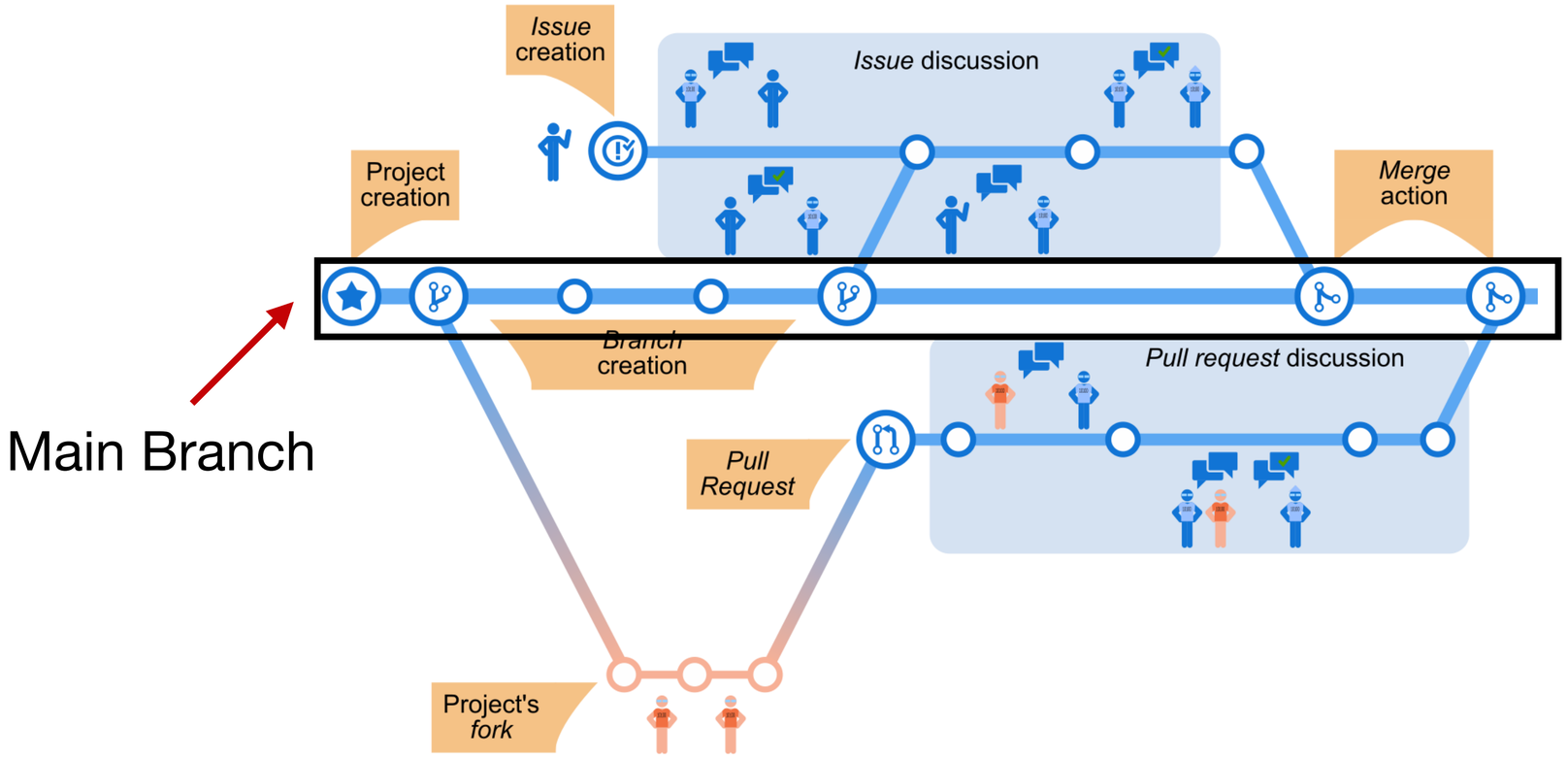
3. **Work Routines** — Core team will transition to admin. role
   (Child and Keiser. 1981, Hambrick and Crozier. 1985)

4. **Coordination** — Work will become open and decentralized
   (Child and Keiser. 1981, Miller. 1994, Whetten. 1987, Gronn. 2002)

Outline

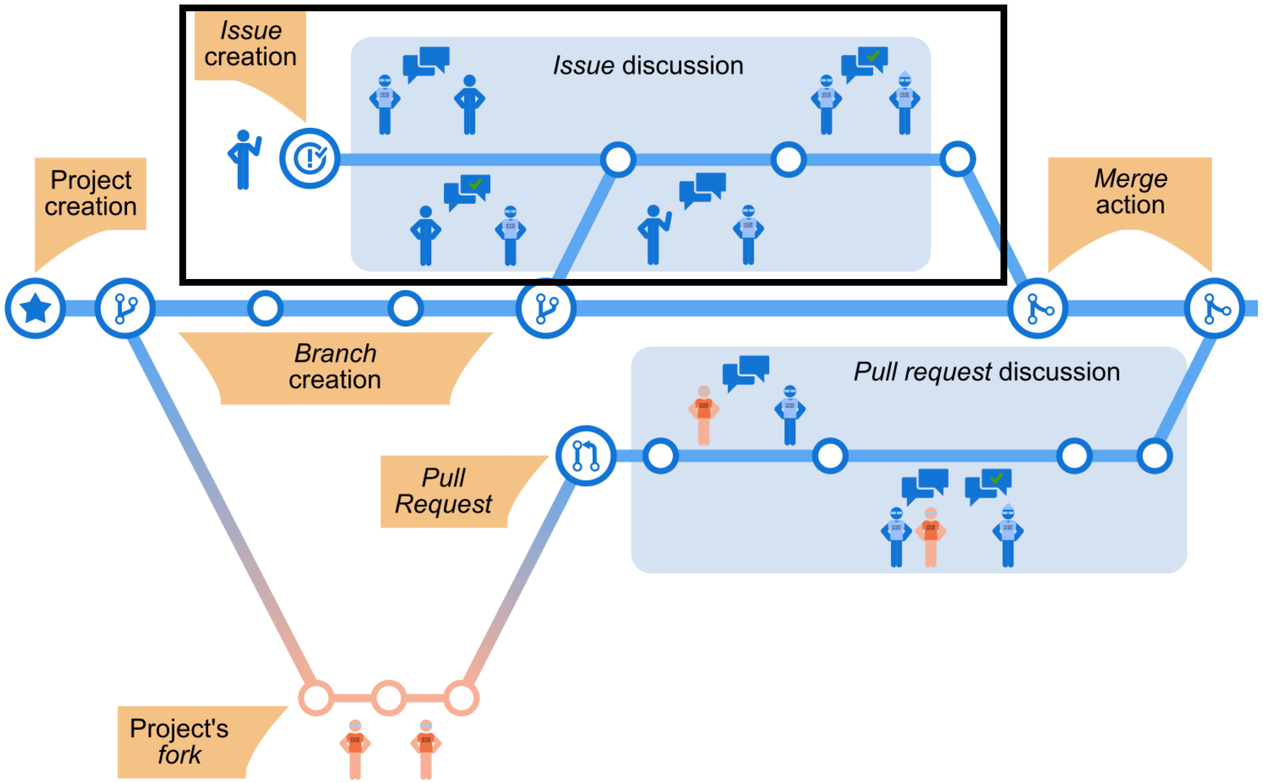
1. Hypotheses
2. **GitHub Workflow**
3. Shocks on GitHub
4. Causal Framework
5. Findings

GitHub Workflow



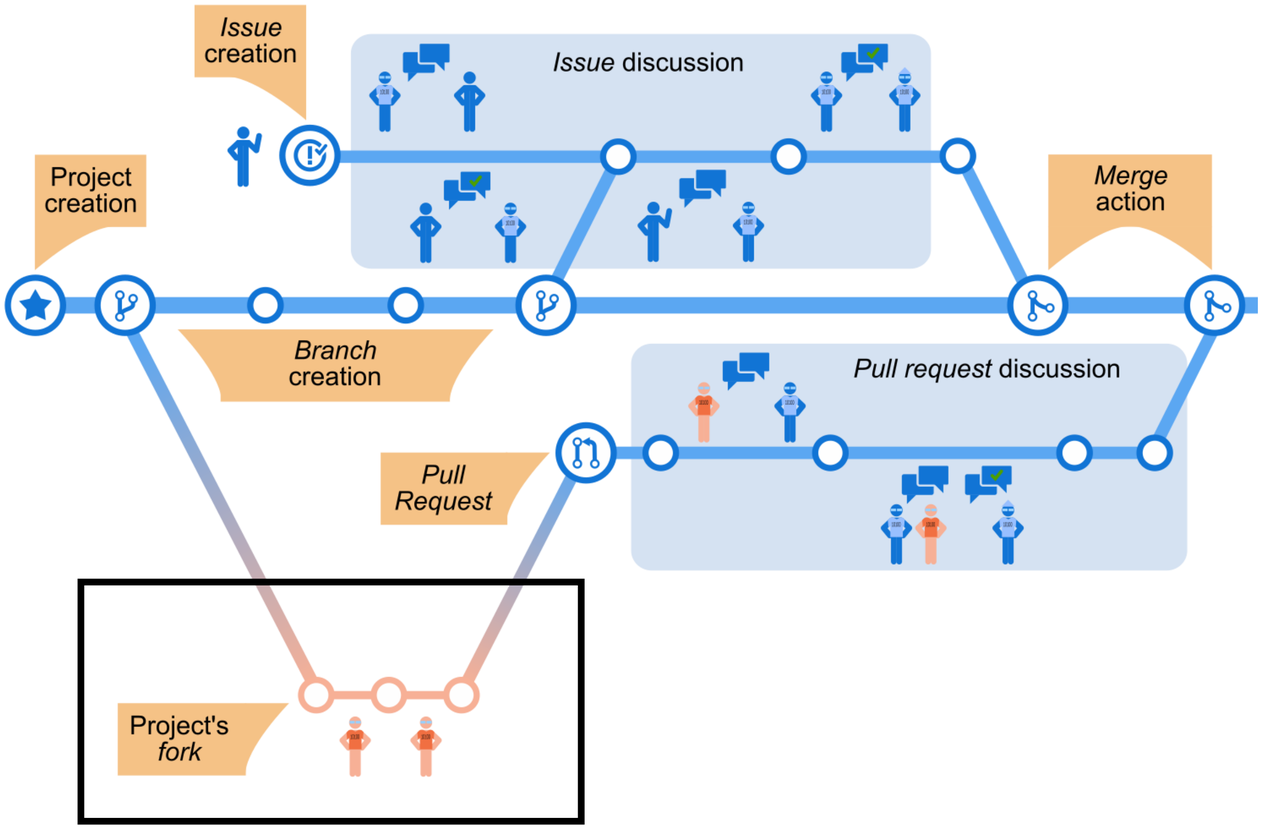
Only members can modify files in the main branch

GitHub Workflow



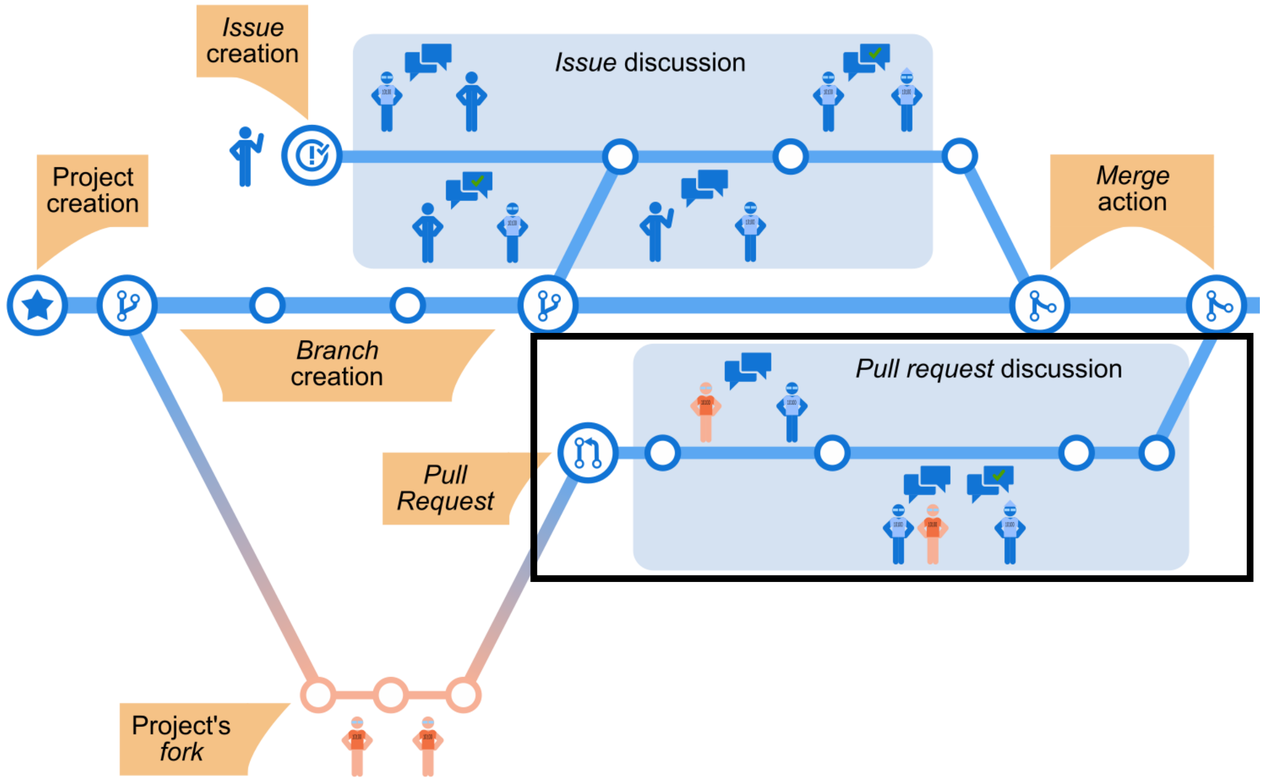
Any user, member or not, can create an issue to discuss a bug, feature request, usage query etc.

GitHub Workflow



Any user can create a copy (fork) of the project for themselves

GitHub Workflow



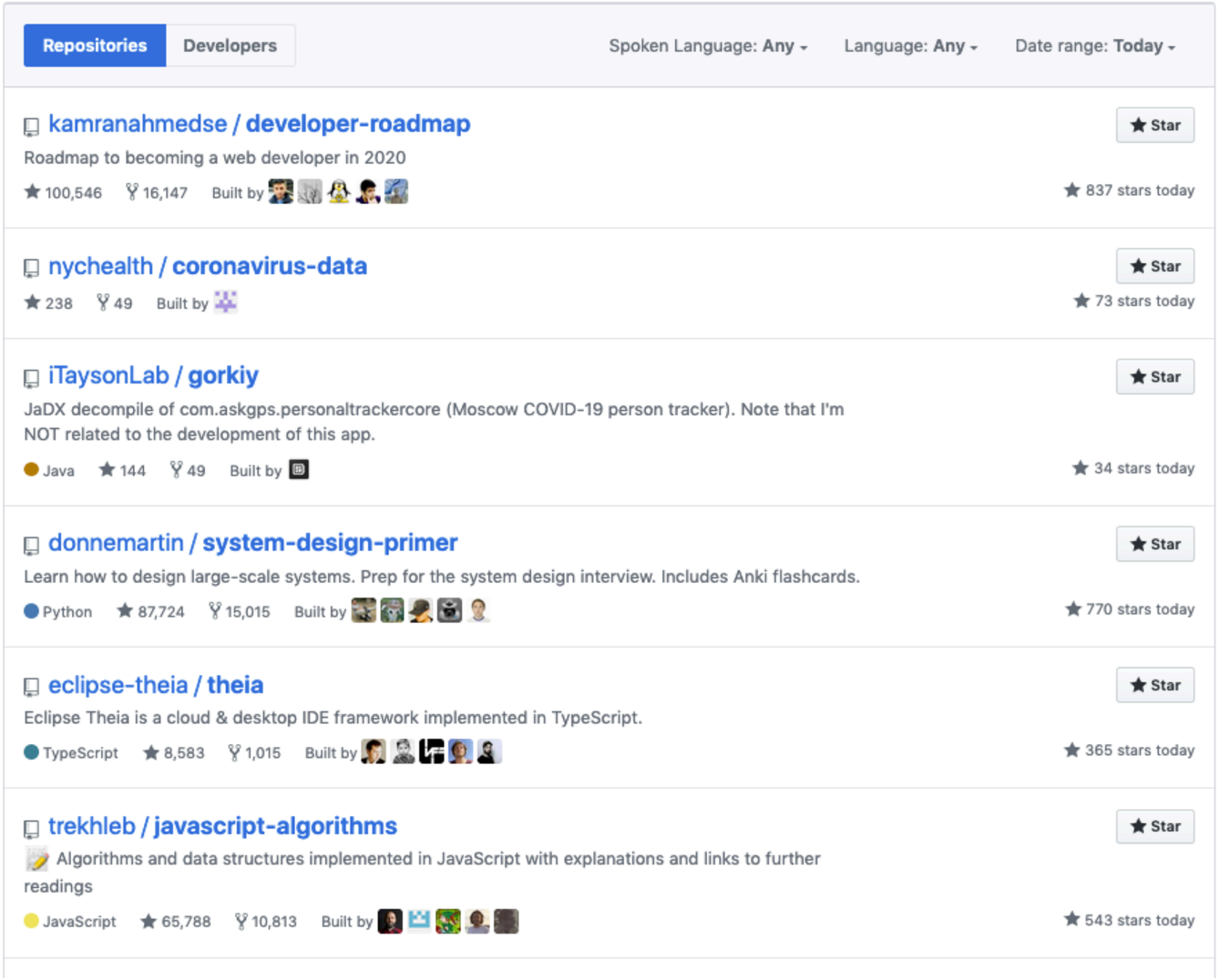
27

External contributors can submit changes made to a fork to the parent project as a pull request — must be approved by members

Outline

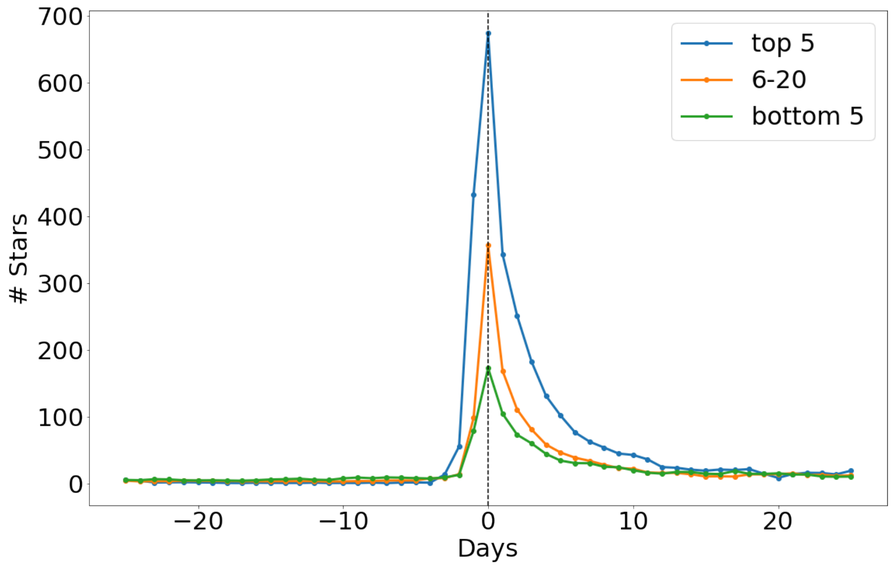
1. Hypotheses
2. GitHub Workflow
3. **Shocks on GitHub**
4. Causal Framework
5. Findings

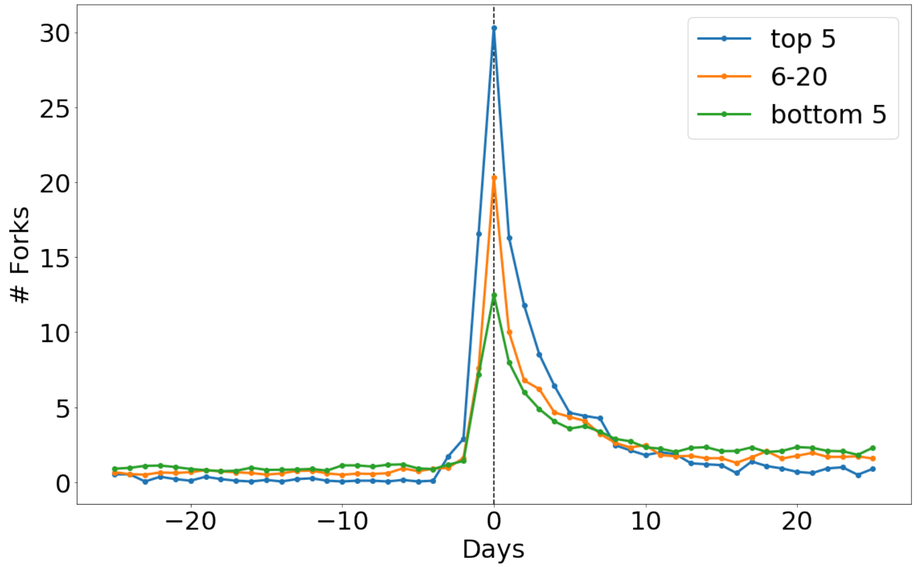
Shocks on GitHub



1. GitHub chooses projects by popularity and activity
2. Updated every 3 hours
3. Overall and language specific lists — 25 projects each
4. Can follow/like projects directly

Trending Drives Community Attention





Day 0 starts from the moment of trending

Shocks Dataset

Trending events
(6/27/18 - 1/31/19)

Shocks Dataset

Trending events
(6/27/18 - 1/31/19)

Remove events during first 14 days

Shocks Dataset

Trending events
(6/27/18 - 1/31/19)

Remove events during first 14 days

Select first best “Top 5” for each project

Shocks Dataset

Trending events
(6/27/18 - 1/31/19)

Remove events during first 14 days

Select first best “Top 5” for each project

Remove projects with no activity before shock

Shocks Dataset

Trending events
(6/27/18 - 1/31/19)

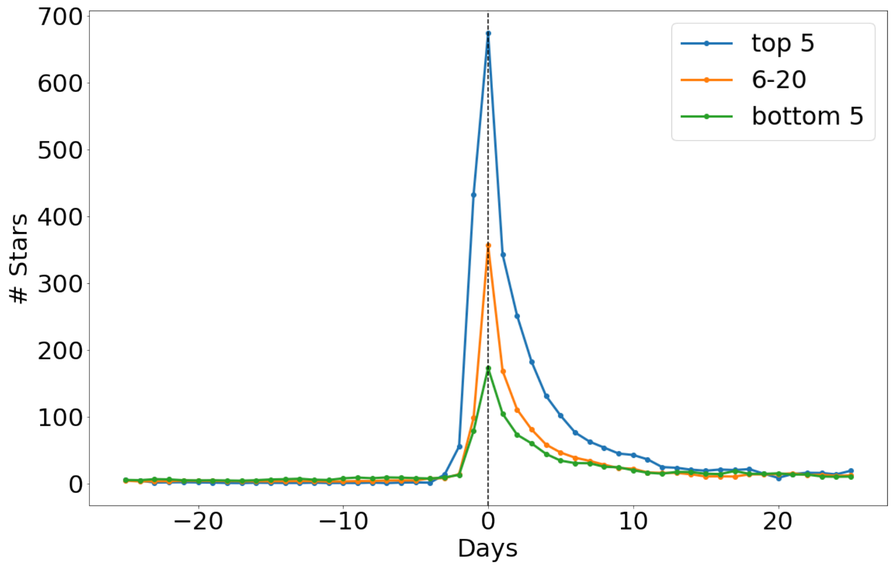
Remove events during first 14 days

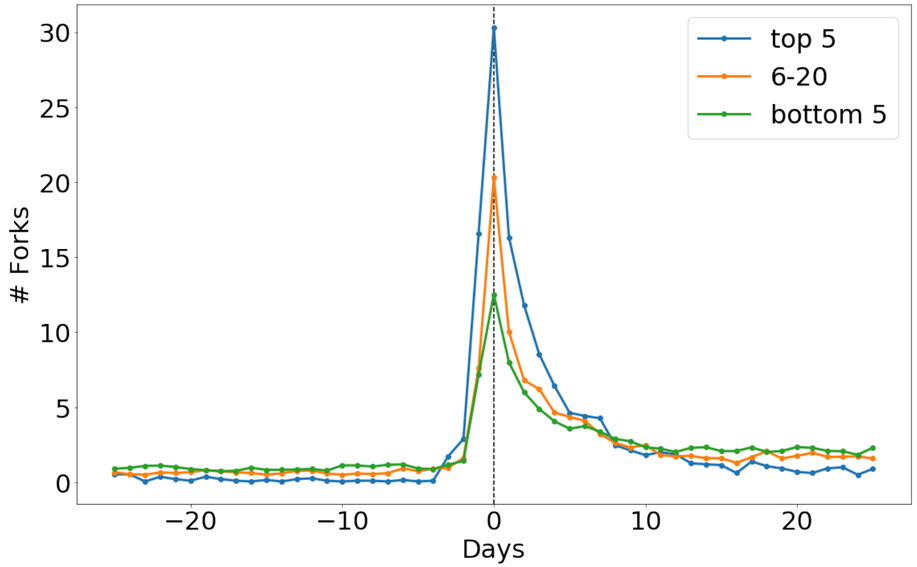
Select first best “Top 5” for each project

Remove projects with no activity before shock

1107 shocks
(project, date, programming language)

Trending Drives Community Attention

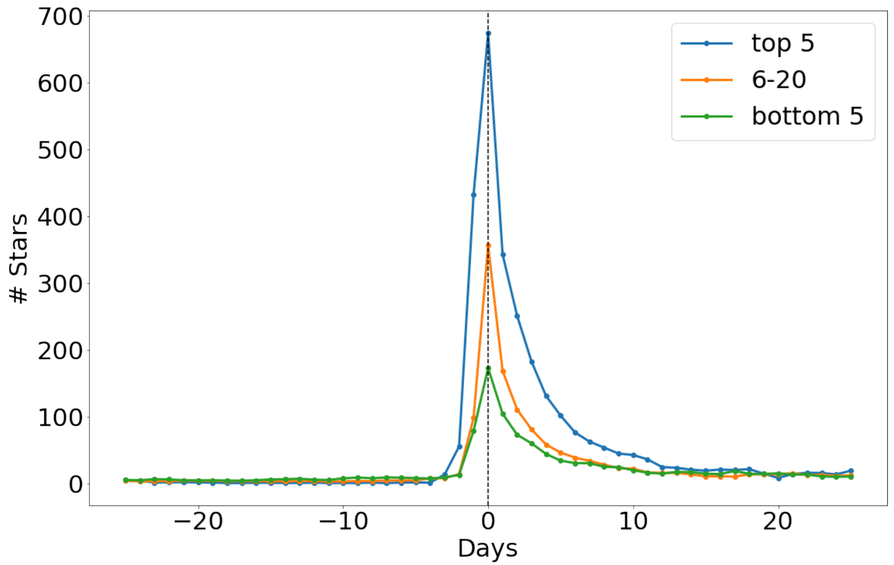


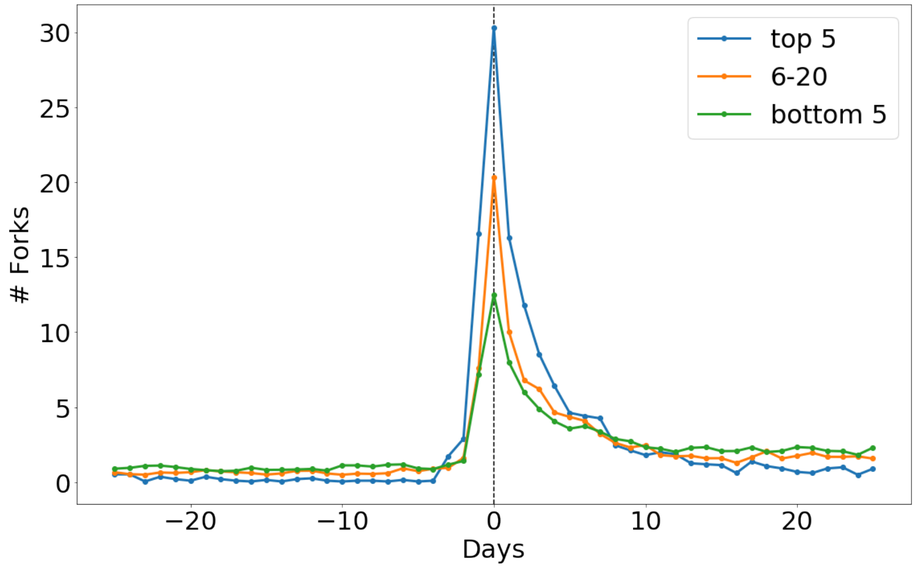


Day 0 starts from the moment of trending

But there's a problem

Trending Drives Community Attention





Day 0 starts from the moment of trending

Projects that trended were already growing and popular before trending

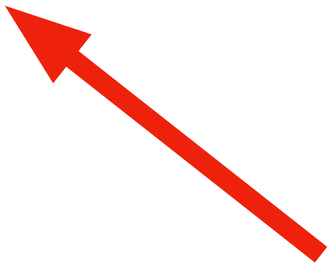
Outline

1. Hypotheses
2. GitHub Workflow
3. Shocks on GitHub
4. **Causal Framework**
5. Findings

Our Approach

Propensity Score Matching + Difference-in-Difference

Identify a control set to account for prior growth



Estimate effect of trending on behavior

Propensity Score Matching

The GitHub trending algorithm primarily uses the numbers of stars and forks and their growth over time to select projects

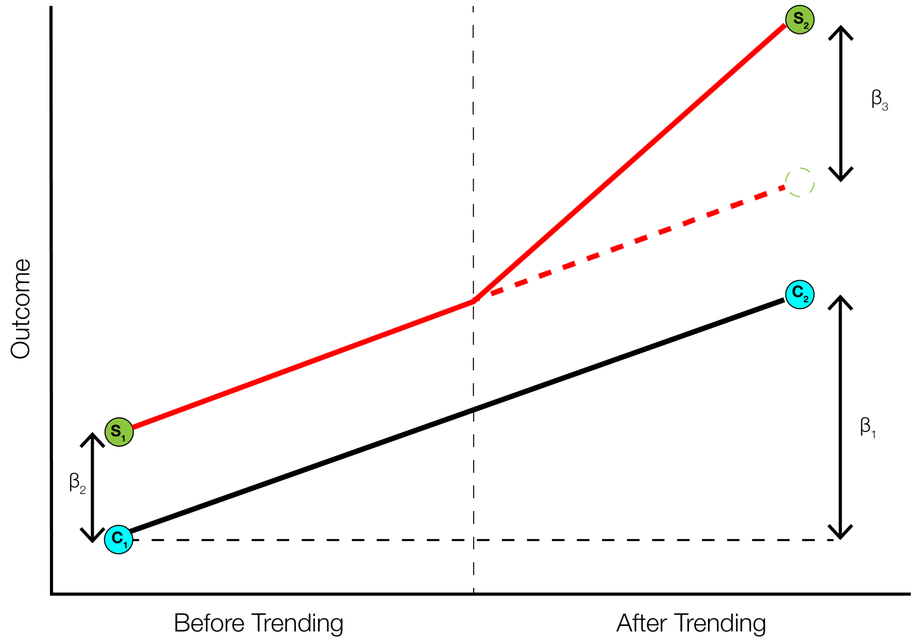
(The GitHub Blog. 2013. Explore what is Trending on GitHub.)

Propensity Score Matching*

1. Estimate propensity of an active project trending on a given date
2. Find nearest neighbors for shocks by propensity and covariate similarity
3. Stratify the combined set of shocks and controls by propensity
4. Are shocks and controls within each strata adequately similar?

* See the paper for additional details

Difference-in-Difference Model



$$y_{it} = \beta_0 + \beta_1 t + \beta_2 I + \beta_3 tI + \beta_4 B + \beta_x C + \epsilon$$

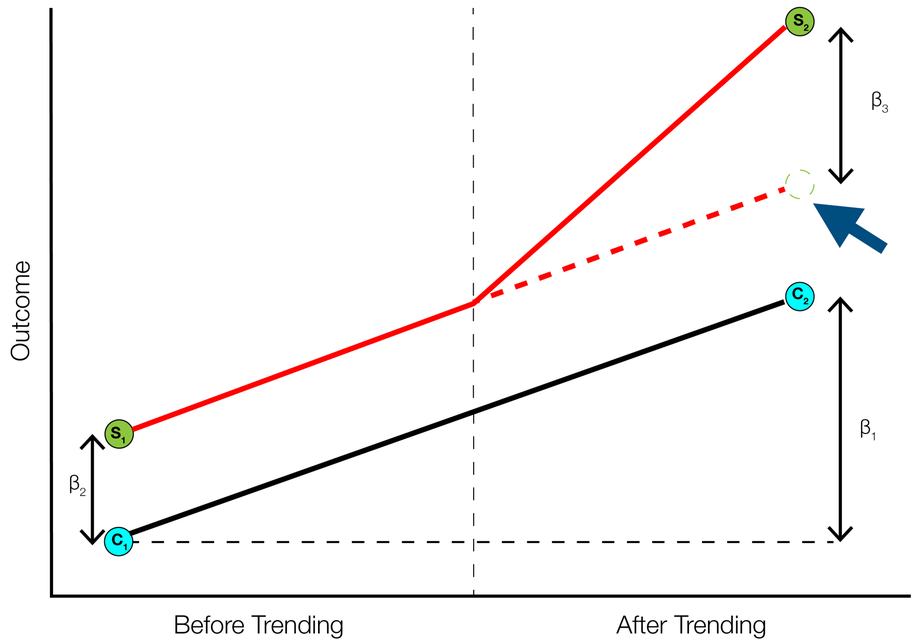
y_{it} - behavior y for project i at period t

$I = 1$, if i is shocked, else 0

B - Fixed effect for PS block

$\beta_x C$ - controls

Difference-in-Difference Model



$$y_{it} = \beta_0 + \beta_1 t + \beta_2 I + \beta_3 tI + \beta_4 B + \beta_x C + \epsilon$$

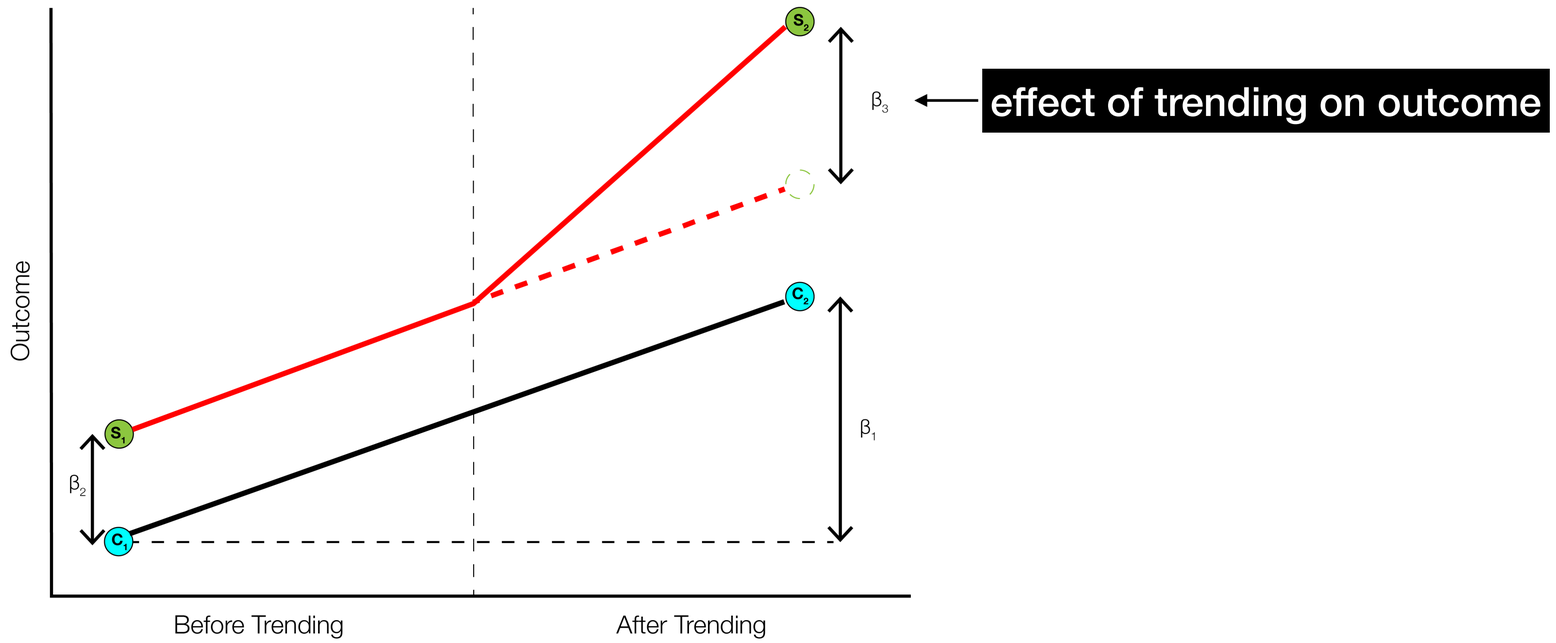
y_{it} - behavior y for project i at period t

$I = 1$, if i is shocked, else 0

B - Fixed effect for PS block

$\beta_x C$ - controls

Difference-in-Difference Model



$$y_{it} = \beta_0 + \beta_1 t + \beta_2 I + \beta_3 tI + \beta_4 B + \beta_x C + \epsilon$$

y_{it} - behavior y for project i at period t

$I = 1$, if i is shocked, else 0

B - Fixed effect for PS block

$\beta_x C$ - controls

Outline

1. Hypotheses
2. GitHub Workflow
3. Shocks on GitHub
4. Causal Framework
5. **Findings**

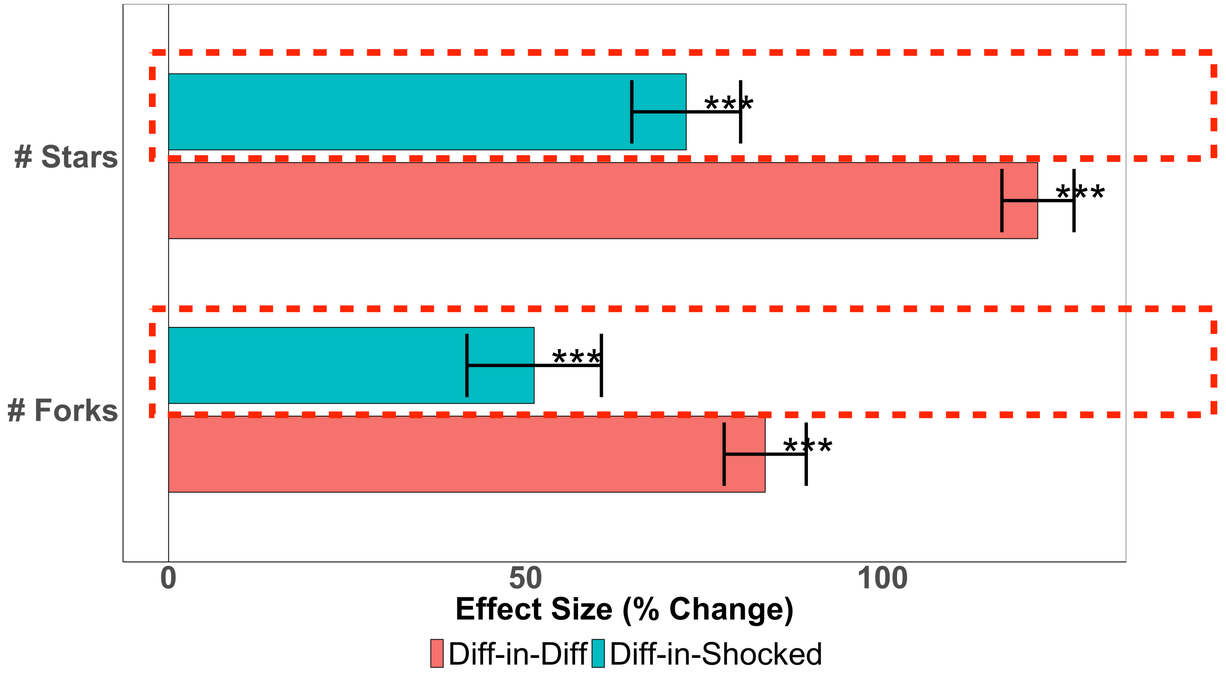
H1: Growth

(a) Community interest will increase

(b) External contributions will increase

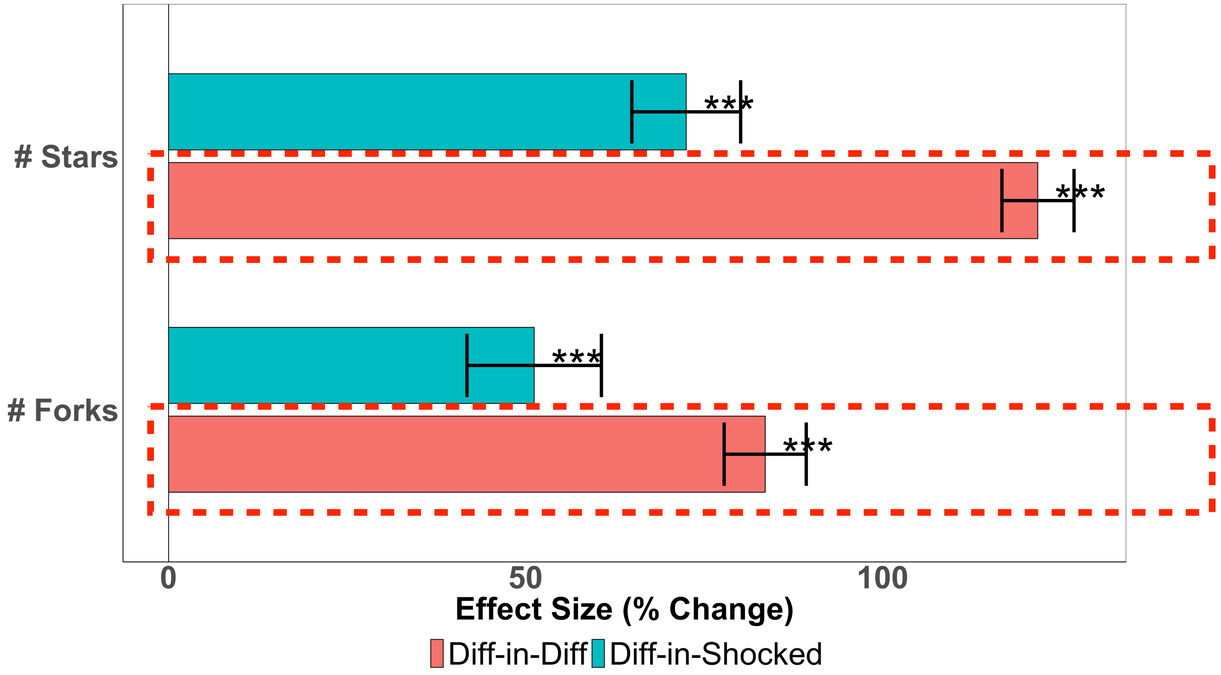
H1: Growth

(a) Community interest will increase



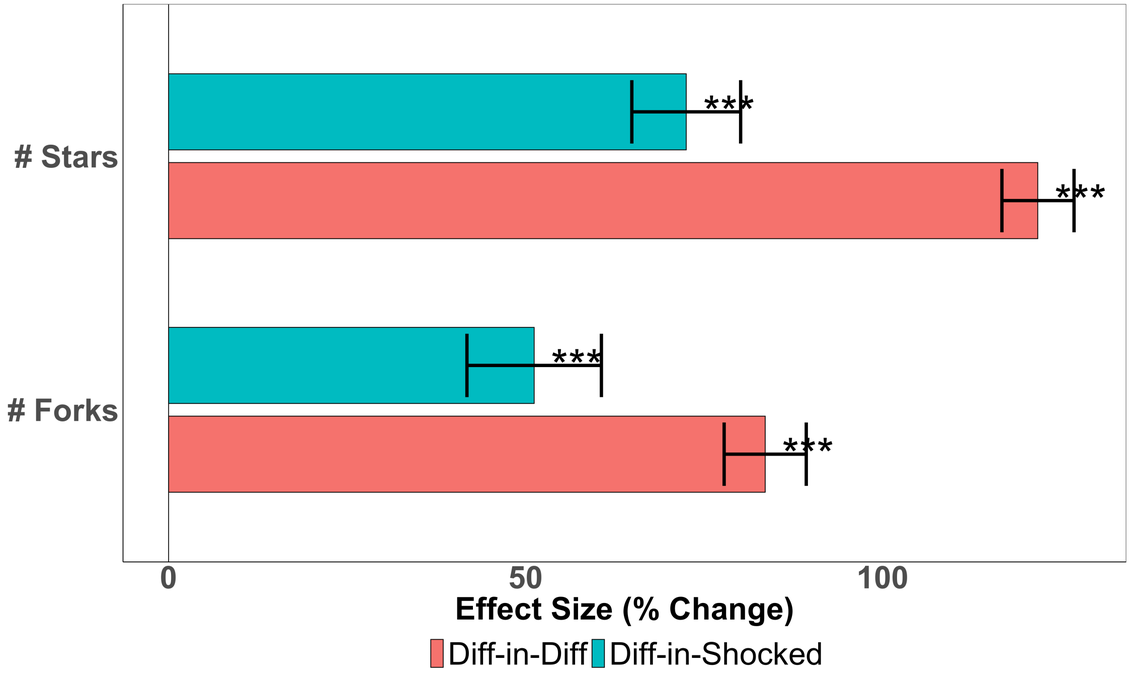
H1: Growth

(a) Community interest will increase



H1: Growth

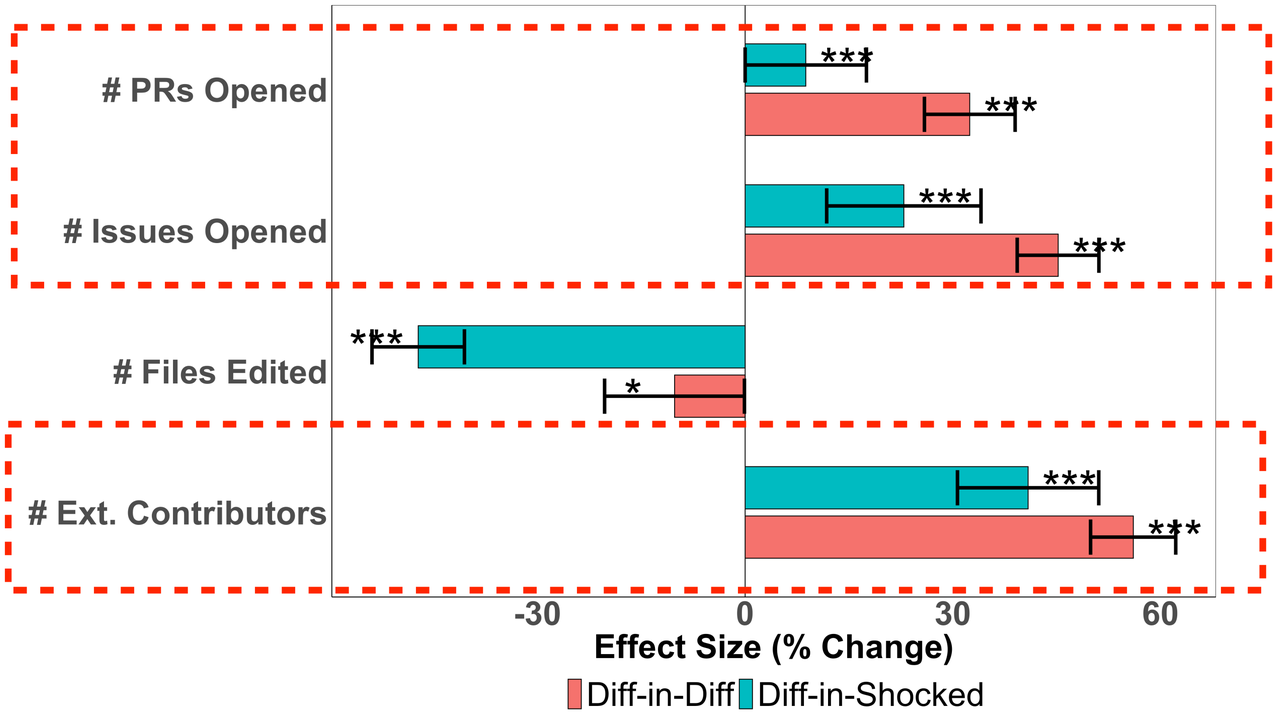
(a) Community interest will increase



The shock leads to an explosion in external attention on a project — while controls lose attention

H1: Growth

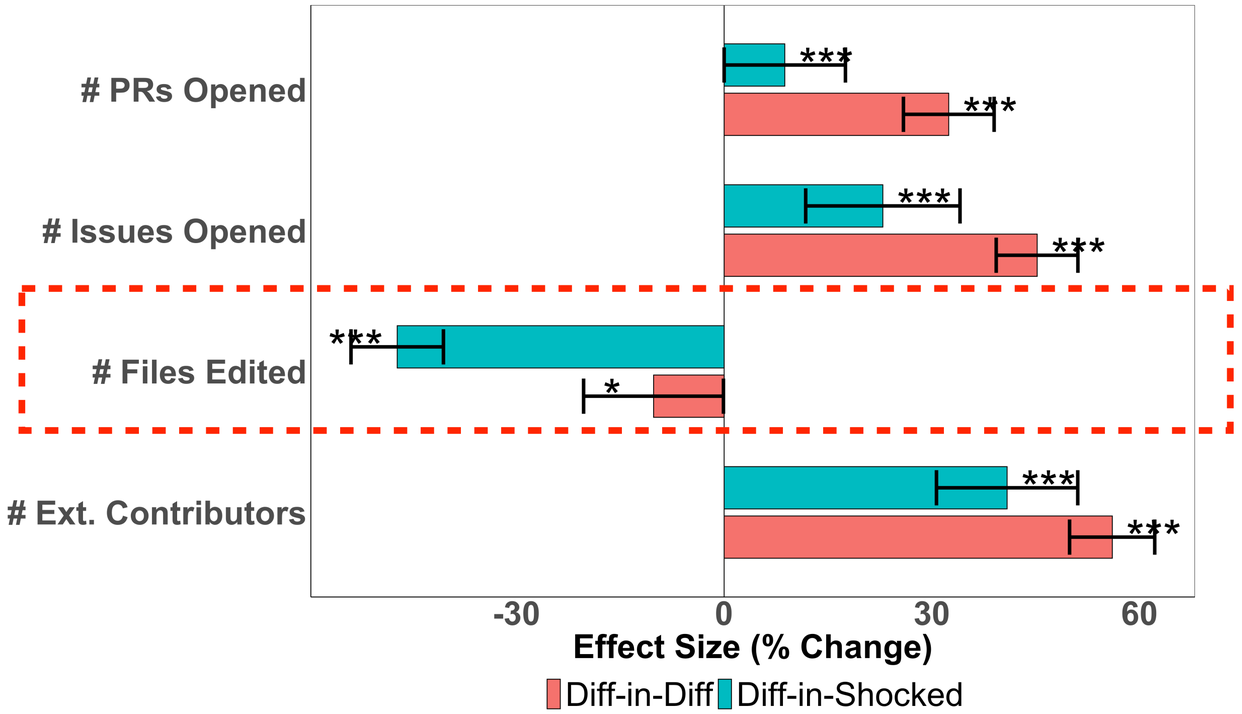
(b) External contributions will increase



Overall contributions also see an increase after the shock

H1: Growth

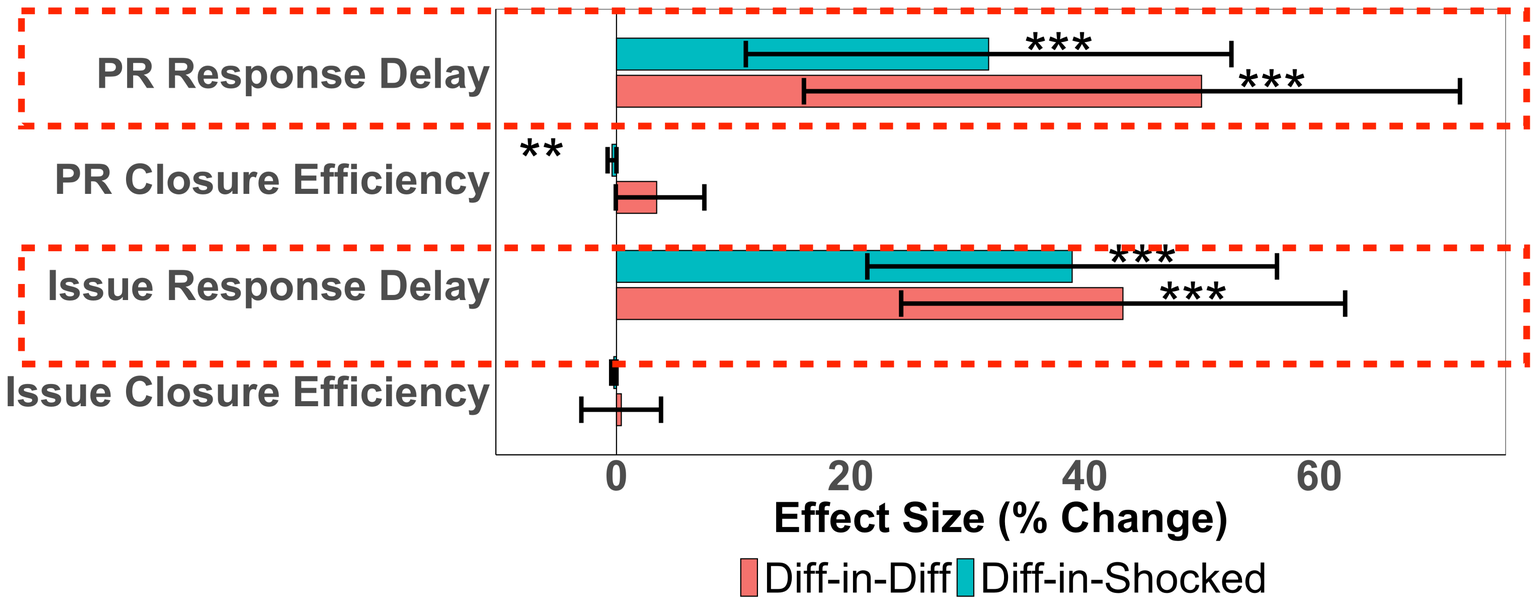
(b) External contributions will increase



But, average work done by an outsider declines — more casual contributions?

H2: Growing Pains

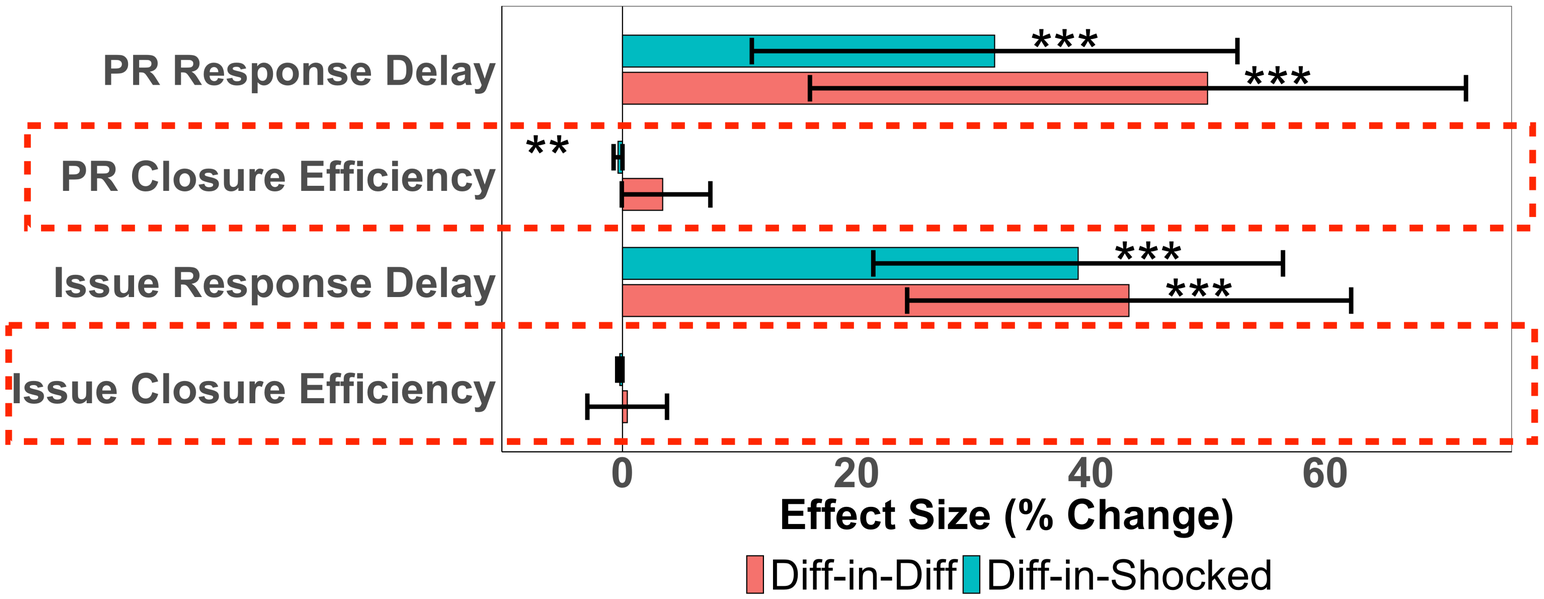
(a) Increased external contributions will lead to larger backlogs with members



Members take longer to respond to *outsiders* after trending

H2: Growing Pains

(b) Increased external contributions will lead to larger backlogs with members



Task (issues/PRs) completion efficiency remains unchanged; may be helped by many more trivial items?

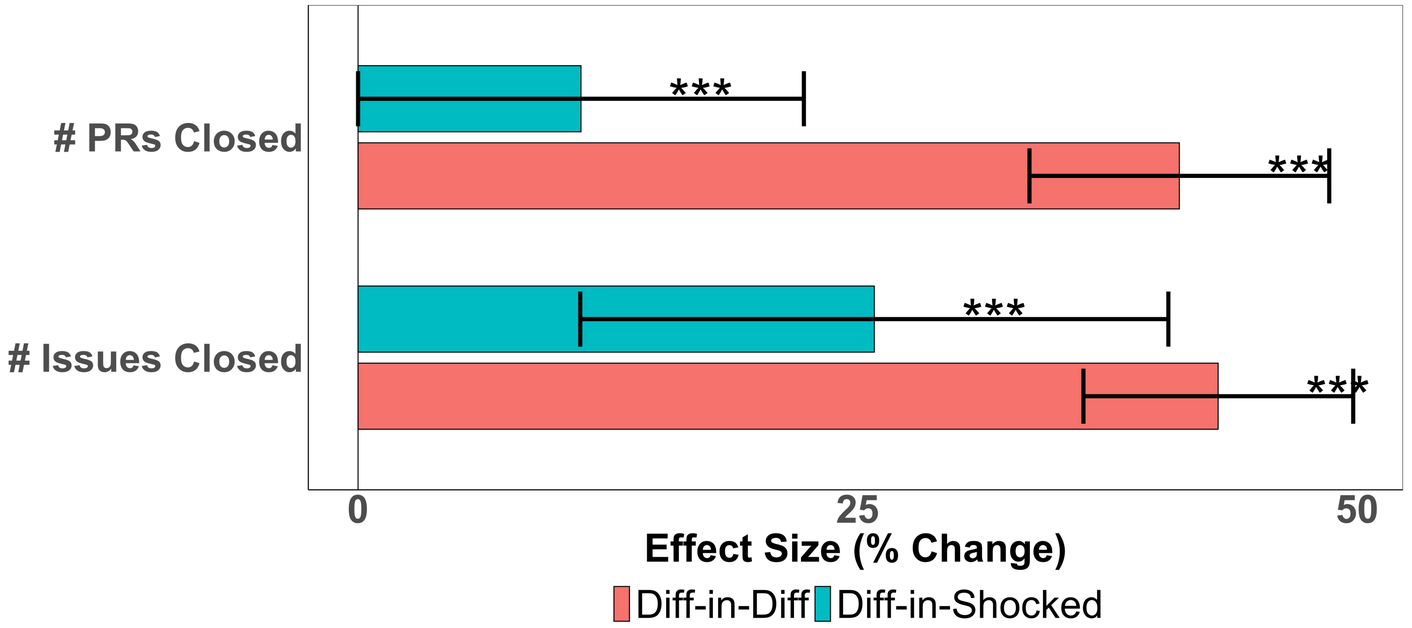
H3: Work Routines

(a) Members will respond to and direct outsiders more

(b) Members will do less development

H3: Work Routines

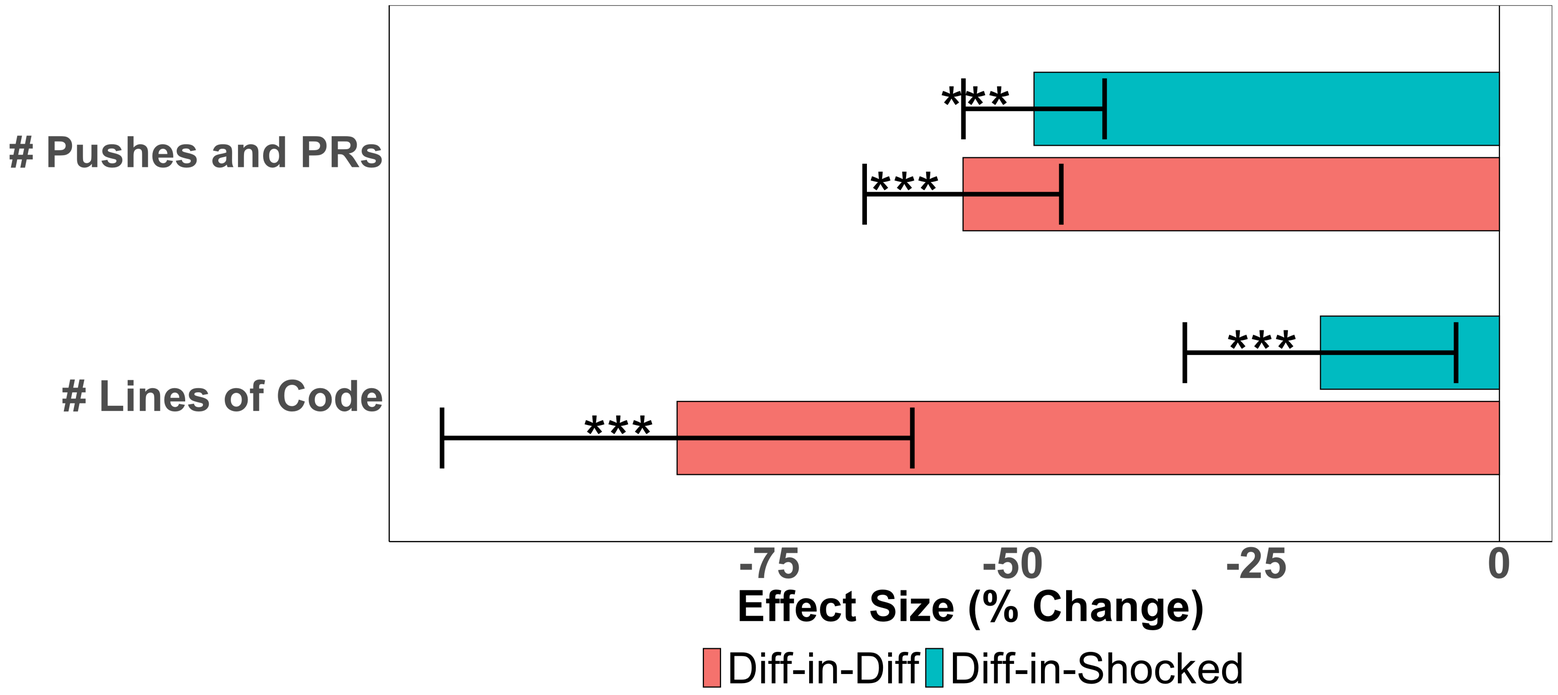
(a) Members will respond to and direct outsiders more



Members are increasing their administrative focus

H3: Work Routines

(b) Members will do less development



Members cut back substantially on coding after trending

H4: Coordination

Distributed Leadership (Gronn, 2002)

(a) Members will increasingly engage with outsiders

H4: Coordination

Distributed Leadership (Gronn, 2002)

(a) Members will increasingly engage with outsiders

(b) Outsiders will become more important within work routines

H4: Coordination

Distributed Leadership (Gronn, 2002)

(a) Members will increasingly engage with outsiders

(b) Outsiders will become more important within work routines

(c) Collaboration will become more modular

H4: Coordination

Distributed Leadership (Gronn, 2002)

(a) Members will increasingly engage with outsiders

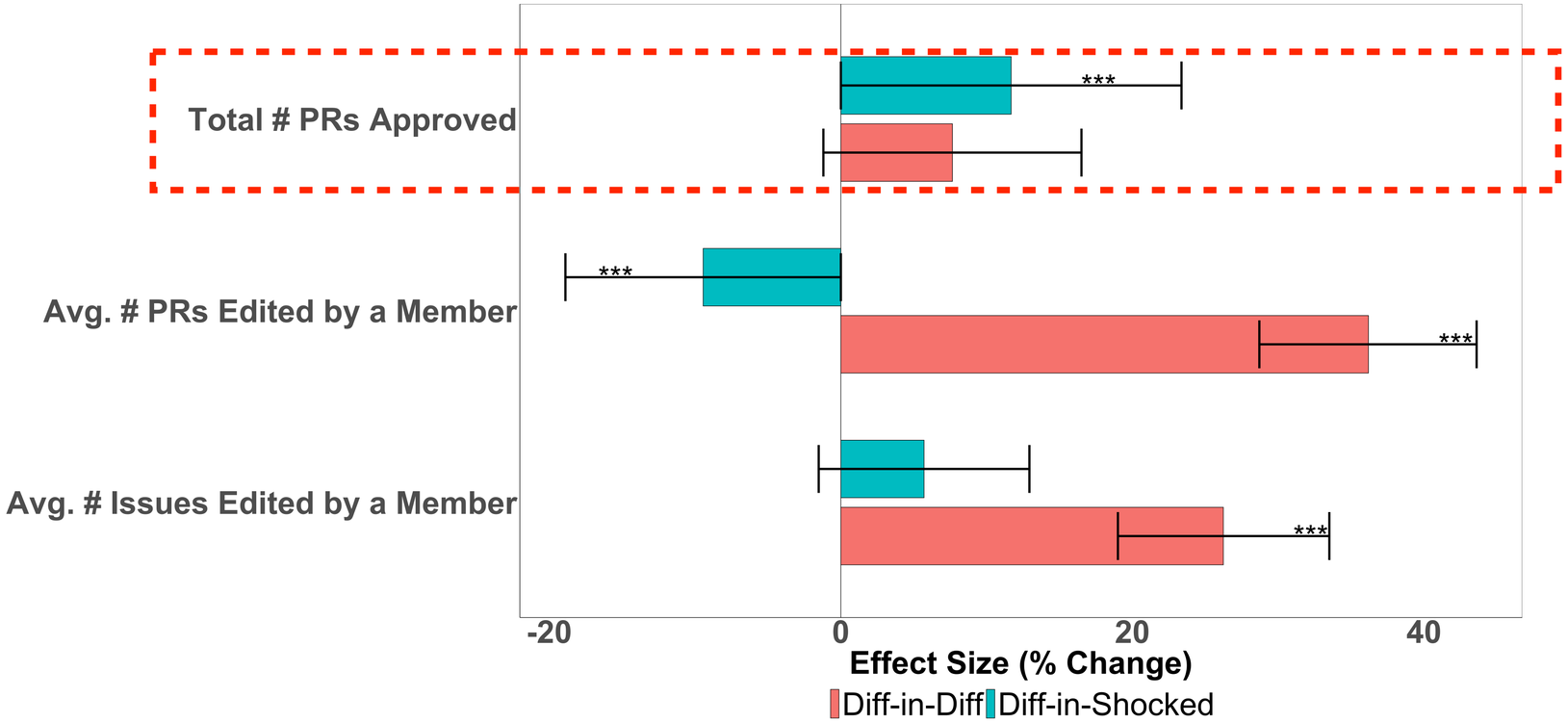
(b) Outsiders will become more important within work routines

(c) Collaboration will become more modular

(d) Members will reinforce core values through automation

H4: Coordination

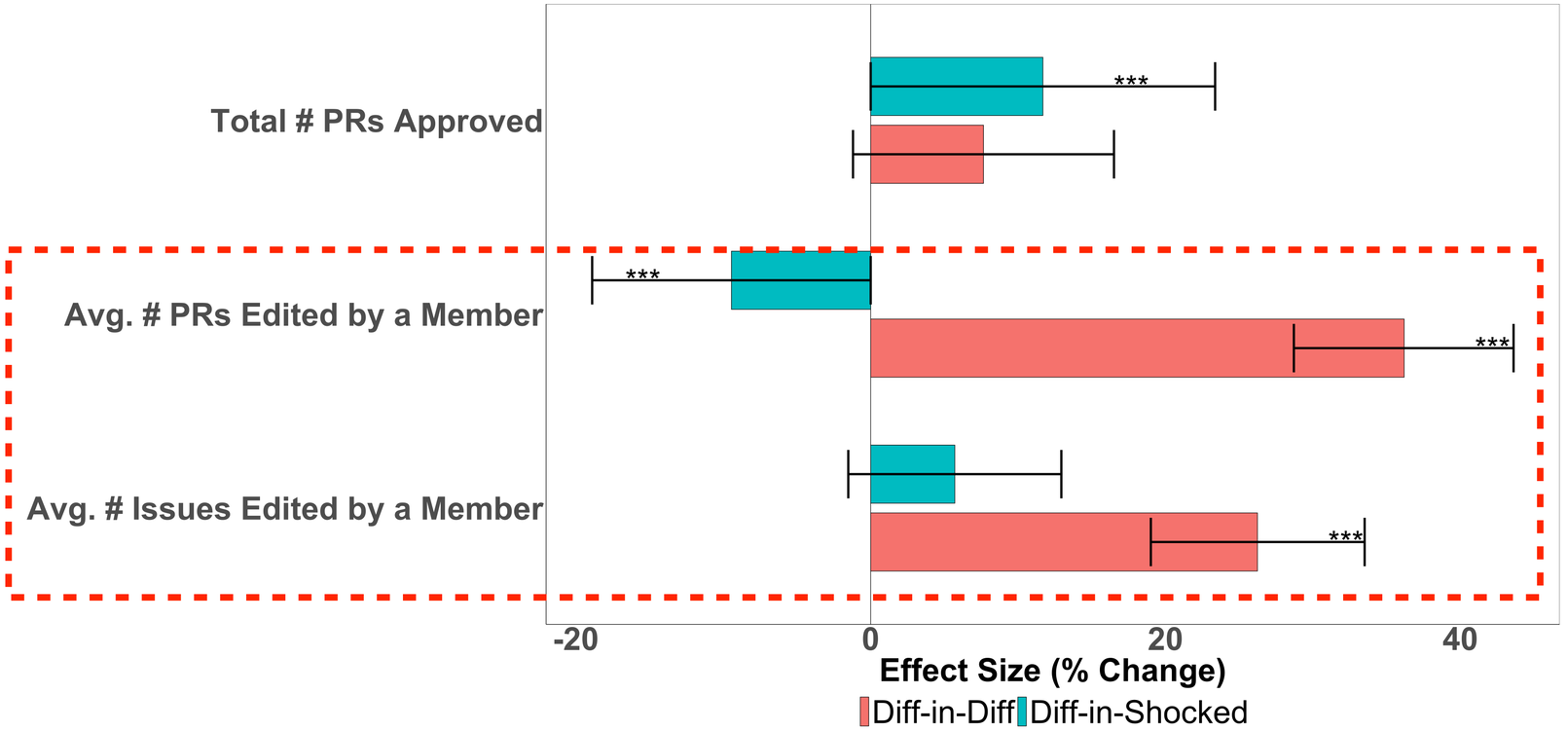
(a) Members will increasingly engage with outsiders



Members collaboratively engage outsiders more through Issue/PR discussions — they also accept more contributions

H4: Coordination

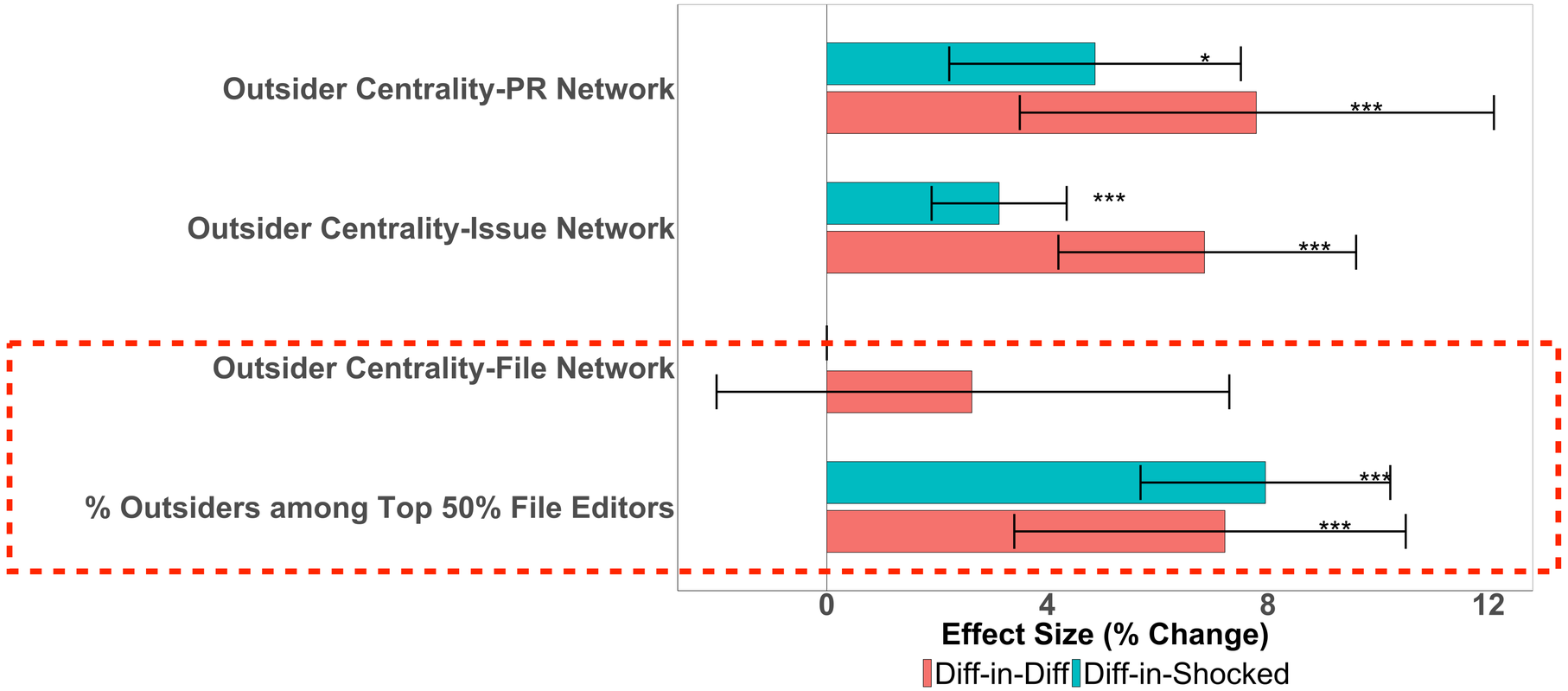
(a) Members will increasingly engage with outsiders



Members collaboratively engage outsiders more through Issue/PR discussions — they also accept more contributions

H4: Coordination

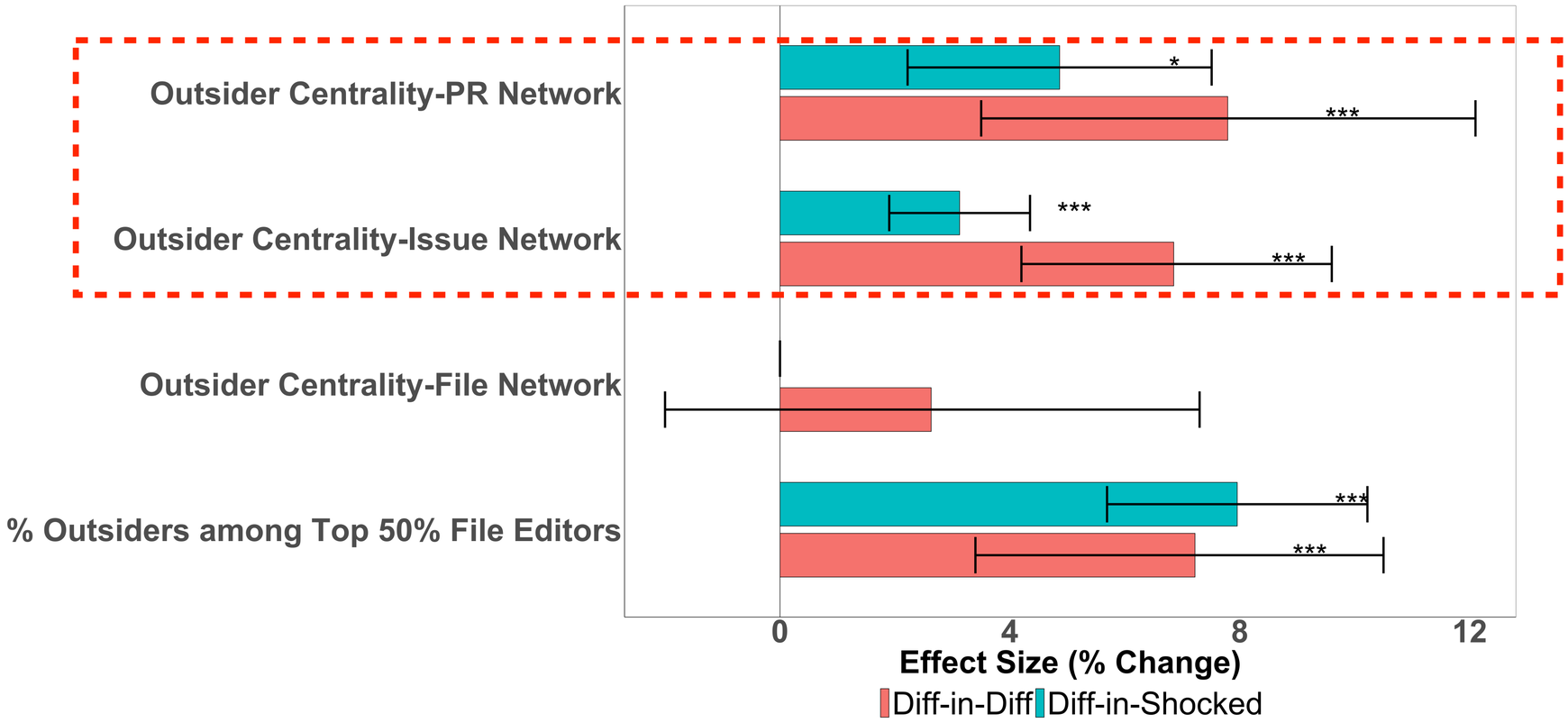
(b) Outsiders will become more important within work routines



Outsiders increase their share of code contributions — but their importance to collaborations doesn't improve

H4: Coordination

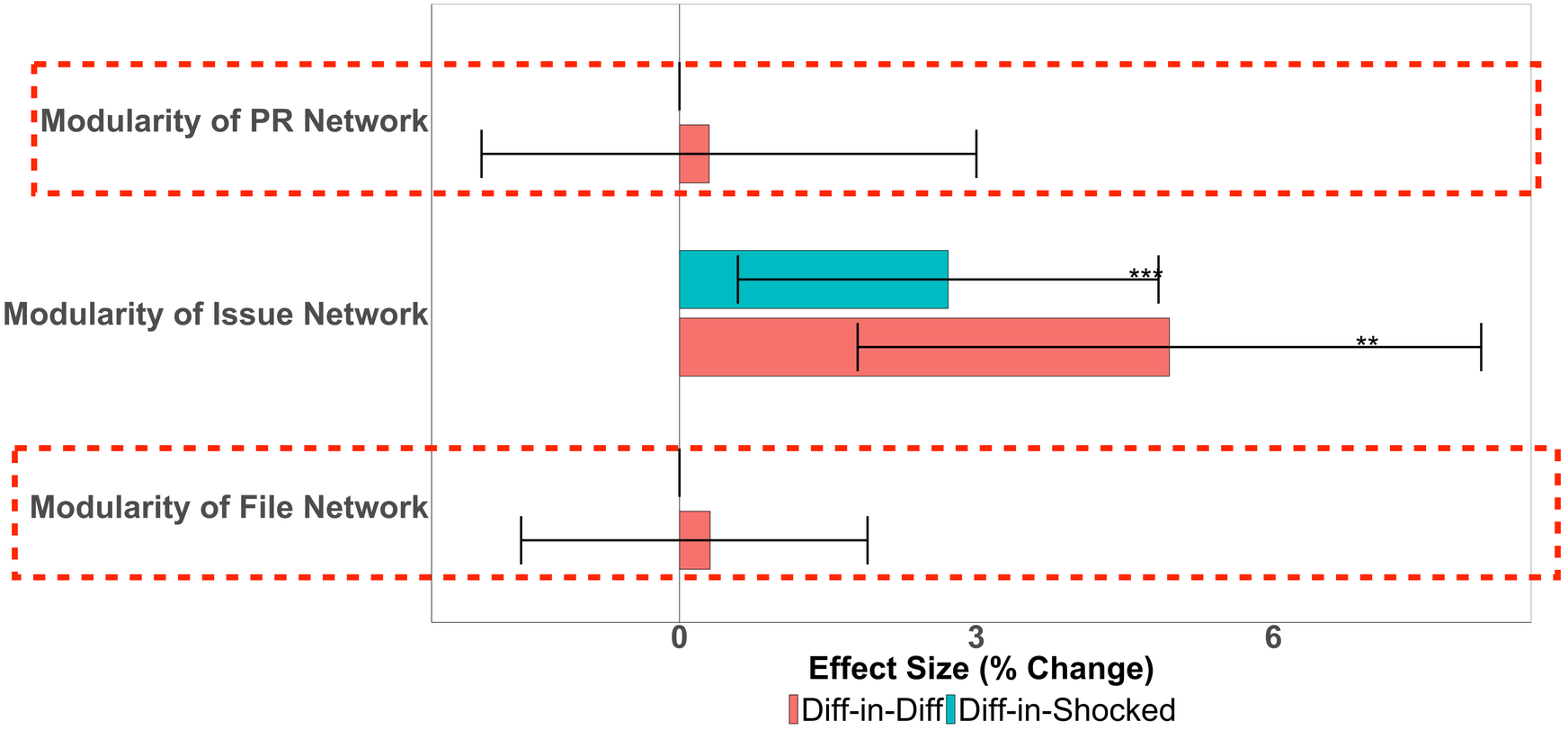
(b) Outsiders will become more important within work routines



Outsiders are clearly becoming more active and important in issue/PR discussions

H4: Coordination

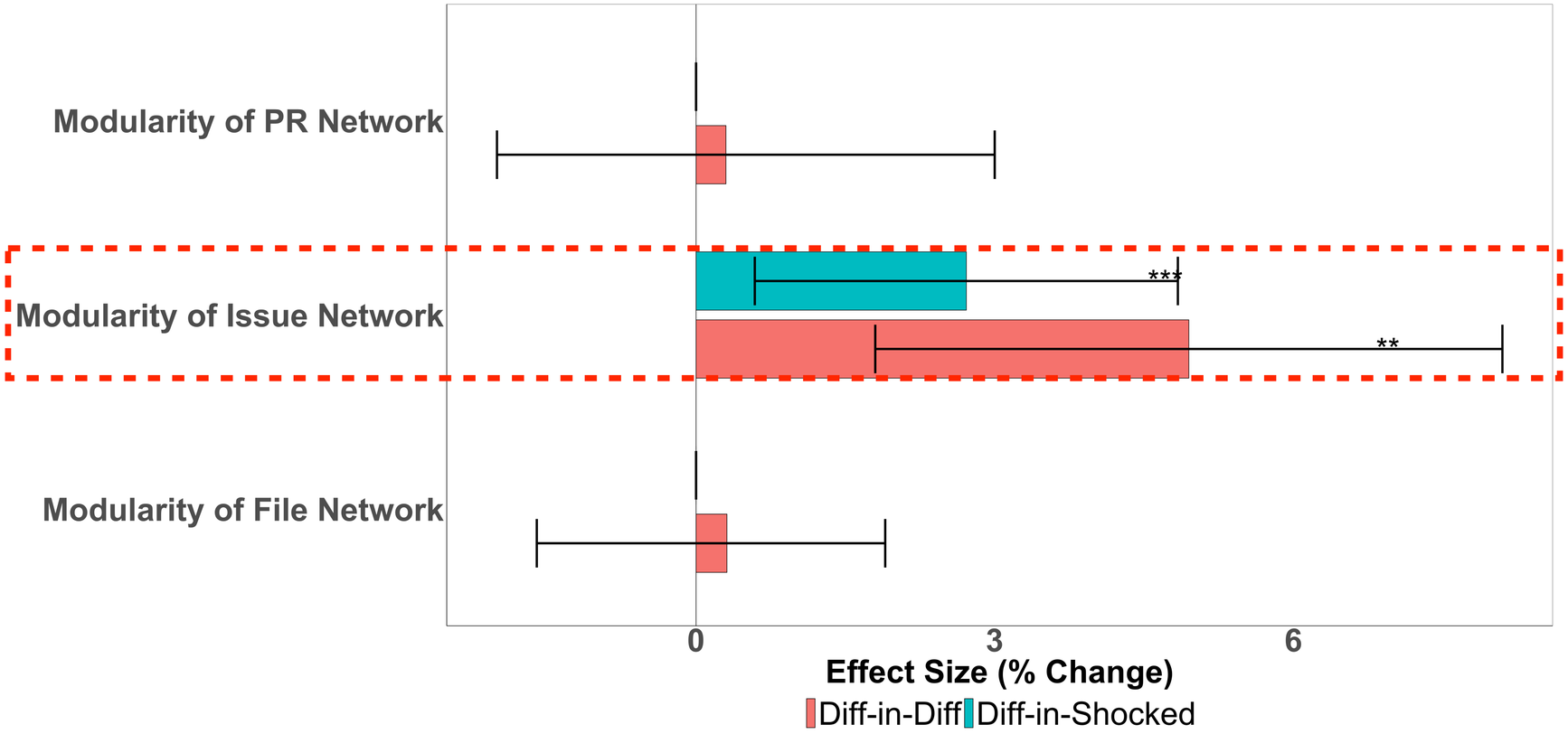
(c) Collaboration will become more modular



Collaboration on the low commitment tasks (issue discussion) becomes slightly more modular

H4: Coordination

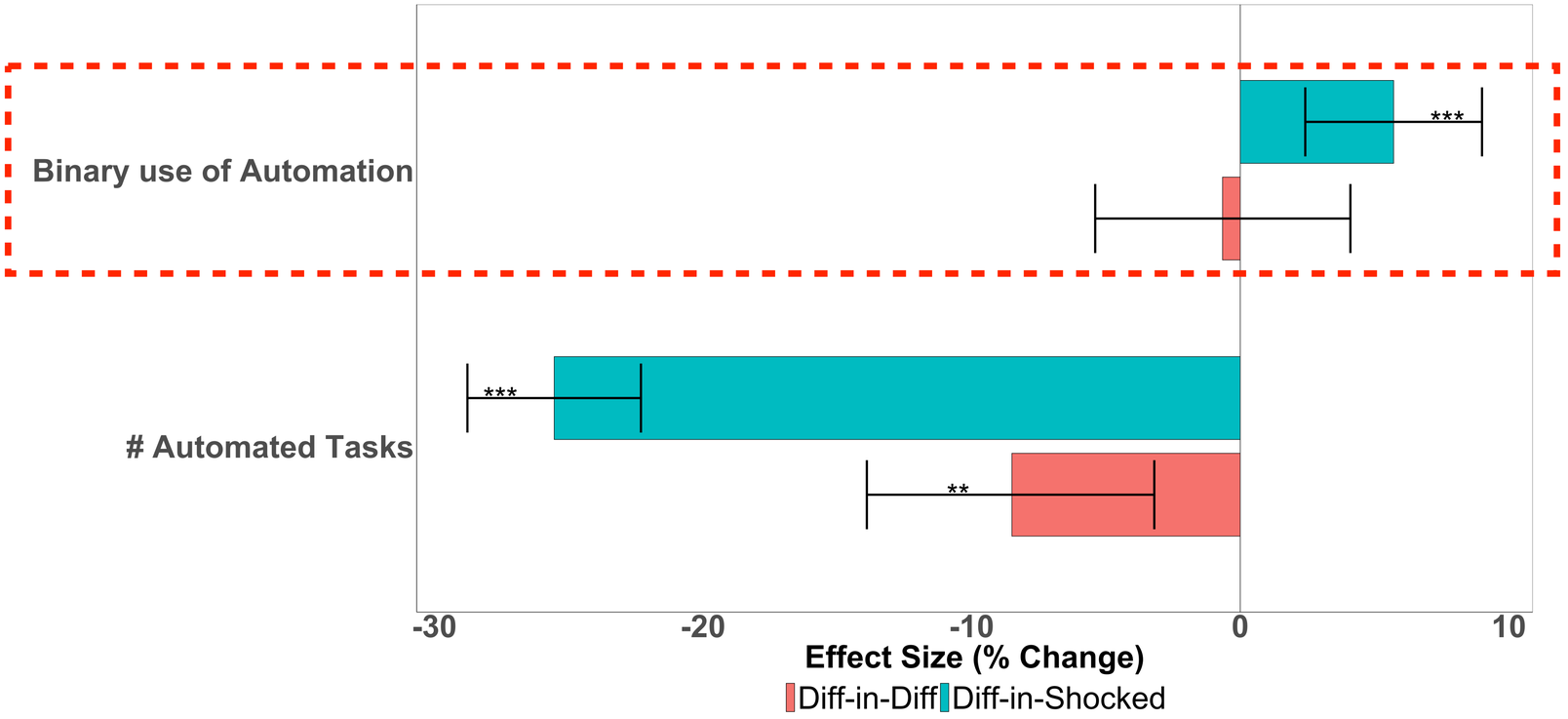
(c) Collaboration will become more modular



Collaboration on the low commitment tasks (issue discussion) becomes slightly more modular

H4: Coordination

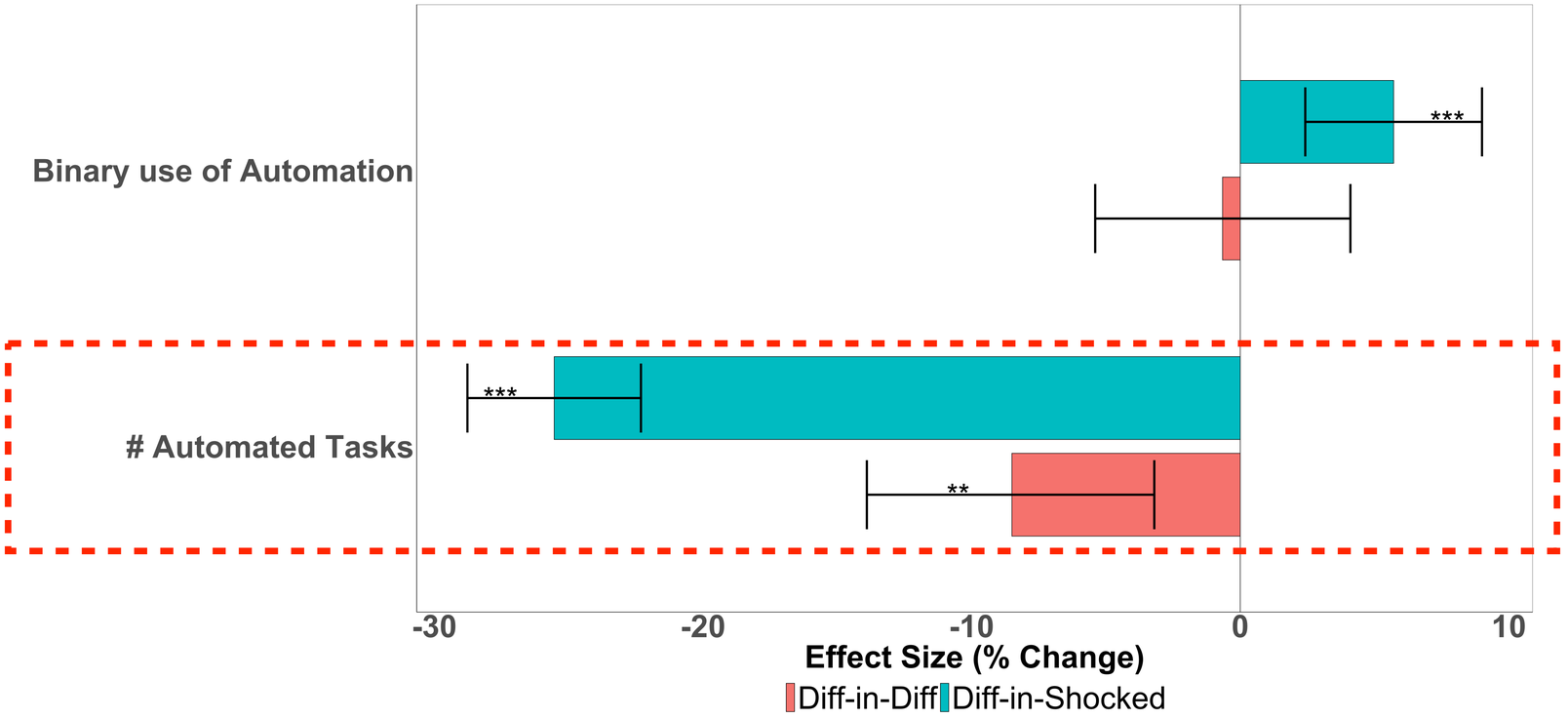
(d) Members will reinforce core values through automation



Projects are more likely to use automation after trending, but no more than the control

H4: Coordination

(d) Members will reinforce core values through automation



Projects use less automated tasks after trending (Potential selection bias?)

Summary

- In the aftermath of an attention shock,
  - Crowds may grow rapidly in size and work done
  - Most new engagements may be shallow
  - The core team will struggle to stay responsive
  - Outsiders will take on more responsibility

Summary

- In the aftermath of an attention shock,
  - Crowds may grow rapidly in size and work done
  - Most new engagements may be shallow
  - The core team will struggle to stay responsive
  - Outsiders will take on more responsibility
- Implications for GitHub teams
  - Dealing with a multitude of shallow contributions may be inefficient for members
  - Low responsiveness towards outsiders may increase newcomer attrition
  - Rapidly changing composition of the core workforce will make maintaining project quality and values challenging.

Summary

- In the aftermath of an attention shock,
  - Crowds may grow rapidly in size and work done
  - Most new engagements may be shallow
  - The core team will struggle to stay responsive
  - Outsiders will take on more responsibility
- Implications for GitHub teams
  - Dealing with a multitude of shallow contributions may be inefficient for members
  - Low responsiveness towards outsiders may increase newcomer attrition
  - Rapidly changing composition of the core workforce will make maintaining project quality and values challenging.
- Next Steps — How do observed behavioral changes affect future performance of projects?