StackOverflow and GitHub: Associations Between Software Development and Crowdsourced Knowledge

Bogdan Vasilescu    Vladimir Filkov    Alexander Serebrenik
TU Eindhoven                    UC Davis                            TU Eindhoven
@b_vasilescu                                                              @aserebrenik
Standing on the shoulders of others

Developers:

• reuse components and libraries

• forage on the Web for information
Standing on the shoulders of others

Developers:

• reuse components and libraries
• forage on the Web for information
Standing on the shoulders of others

Developers:

- reuse components and libraries
- forage on the Web for information
How do I edit an incorrect commit message in Git?

I stupidly did a Git commit while half asleep, and wrote the totally wrong thing in the commit message. How do I change the commit message?

I have not yet pushed the commit to anyone.

For those somewhat new to git: Laurie's point about having not yet pushed is important. Like rebasing, this is changing the history. If someone has cloned/pulled from your repo between the original and

20 Answers

```
$ git commit --amend -m "New commit message"
```

Used to amend the tip of the current branch. Prepare the tree object you would want to replace the latest commit as usual (this includes the usual -i/-o and explicit paths), and the commit log editor is seeded with the commit message from the tip of the current branch. The commit you create replaces the current tip -- if it was a merge, it will have the parents of the current tip as parents -- so the current top commit is discarded.

It is a rough equivalent for:

```
$ git reset --soft HEAD
$ ... do something else to come up with the right tree ...
$ git commit -c ORIG_HEAD
```

but can be used to amend a merge commit.
Writing code vs. seeking and sharing knowledge

Demand for knowledge

Supply of knowledge
Is participation in SO related to productivity of developers?
Is participation in SO related to productivity of developers?

Beneficial:

• good technical solutions

  [Parnin et al. “Crowd documentation: Exploring the coverage and the dynamics of API discussions on Stack Overflow,” Georgia Institute of Technology, Tech. Rep., 2012]

• fast answers (median 11 mins)

Is participation in SO related to productivity of developers?

Detrimental:

- competes for time

- gamified, thus addictive

- context switches are expensive
Is participation in SO related to productivity of developers?

Asset or burden?
Dataset

Largest code host in the world

![GitHub](https://github.com)

~400k users
July 2011 - April 2012

Largest programming Q&A site in the world

![Stack Overflow](https://stackoverflow.com)

~1.3M users
July 2008 - August 2012

---


[Quarterly StackExchange data dump (August 2012)]
Dataset

[Quarterly StackExchange data dump (August 2012)]

July 2011 - April 2012

July 2008 - August 2012

Dataset

Email address (plain text)

Email address (MD5 hash)

July 2011 - April 2012

July 2008 - August 2012


[Quarterly StackExchange data dump (August 2012)]
Dataset

Email address (plain text)

Email address (MD5 hash)

July 2011 - April 2012

~94k users

(24%)

(7%)

July 2008 - August 2012


[Quarterly StackExchange data dump (August 2012)]
Email address (plain text)

Email address (MD5 hash)

~47k users active on both GitHub and StackOverflow between July 2011 - April 2012

July 2011 - April 2012

July 2008 - August 2012


[Quarterly StackExchange data dump (August 2012)]
Is participation in SO related to productivity of developers?

Asset or burden?
Macro: overall activity levels

To what extent can activity (expertise) on one platform be used as a proxy for activity (expertise) on the other?

- social signals (e.g., open source projects, professional social media) ~ career advancement

Intermediate: working rhythms

Is attention **focused** (bursts of commits) or **divided** between the two platforms?

- working rhythms of developers ~ software quality

Do StackOverflow activities accelerate or slow down GitHub commits?

<table>
<thead>
<tr>
<th>Name</th>
<th>#Commits</th>
<th>#Questions</th>
<th>#Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dave</td>
<td>100</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Stuart</td>
<td>10</td>
<td>75</td>
<td>15</td>
</tr>
<tr>
<td>Kevin</td>
<td>25</td>
<td>10</td>
<td>75</td>
</tr>
</tbody>
</table>
## Macro

### Overall activity

<table>
<thead>
<tr>
<th></th>
<th>#Commits</th>
<th>#Questions</th>
<th>#Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dave</td>
<td>100</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Kevin</td>
<td>25</td>
<td>10</td>
<td>75</td>
</tr>
<tr>
<td>Stuart</td>
<td>10</td>
<td>75</td>
<td>15</td>
</tr>
</tbody>
</table>
## Macro

### Overall activity

<table>
<thead>
<tr>
<th></th>
<th>#Commits</th>
<th>#Questions</th>
<th>#Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dave</td>
<td>100</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Kevin</td>
<td>25</td>
<td>10</td>
<td>75</td>
</tr>
<tr>
<td>Stuart</td>
<td>10</td>
<td>75</td>
<td>15</td>
</tr>
</tbody>
</table>

Fix, sort

**Not restricted to monotonic relations!**

quartiles/Deciles, compare
Active GitHub committers are experienced developers:
- few StackOverflow questions
- many StackOverflow answers
• Active GitHub committers are experienced developers:
  • few StackOverflow questions
  • many StackOverflow answers

Top StackOverflow users are superstars rather than slackers!
Findings

- Active GitHub committers are experienced developers:

  GitHub activity ~ StackOverflow willingness to answer technical questions (expertise)

  Top StackOverflow users are superstars rather than slackers!
Intermediate Working rhythms

Committing rhythm: series of inter-commit time intervals


Gini (Committing rhythms): focused vs. distributed attention

Gini (Committing rhythms)  

<table>
<thead>
<tr>
<th></th>
<th>#Q</th>
<th>#A</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Q</td>
<td>75</td>
<td>15</td>
</tr>
</tbody>
</table>
Findings

- **Asking** questions on StackOverflow influences how developers distribute their GitHub commits:
  - heavy askers: bursts of intense commit activity followed by longer periods of inactivity (focused attention)
Asking questions on StackOverflow influences how developers distribute their GitHub commits:

- **heavy askers:** bursts of intense commit activity followed by longer periods of inactivity (focused attention)

Learning from StackOverflow (by asking) and committing to GitHub
Who benefits from participating in SO

Dave

Who benefits from participating in SO

Who benefits from participating in SO

Who benefits from participating in SO

Compare actual and shuffled series:

actual < shuffled: acceleration

actual > shuffled: impediment

Findings
Findings

For active committers, asking and answering questions on StackOverflow catalyses committing on GitHub.

For no group is participating in StackOverflow detrimental!
Summary:
Is participation in SO related to productivity of GitHub dev’s?
Asset or burden?
Experts are experts everywhere!

Active committers are also active answerers (knowledge providers) on stackoverflow.

Different working rhythms for novices (focused attention) and experts.

Participating in stackoverflow reinforces commit activities on github.

Going to stackoverflow is “costlier” for novices.

Asset or burden.
Summary

Active [github](https://github.com) committers are also active answerers (knowledge providers) on [stackoverflow](https://stackoverflow.com).

Different working rhythms for novices (focused attention) and experts.

Participating in [stackoverflow](https://stackoverflow.com) reinforces commit activities on [github](https://github.com).

Experts are experts everywhere!

Going to [stackoverflow](https://stackoverflow.com) is “costlier” for novices.

[asset](https://example.com) or [burden](https://example.com)