Going Farther Together: The Impact of Social Capital on Sustained Participation in Open Source

Huilian Sophie Qiu, Alexander Nolte, Anita Brown, Alexander Serebrenik, and Bogdan Vasilescu

@sophiehsqq @alexander_nolte @aserebrenik @b_vasilescu
Maintain, or suffer

https://w3techs.com/technologies/history_overview/web_server

(Greenstein and Nagel, 2016)
Open-source experience boosts resumes

Employers (and job seekers) use open-source experience to make inferences (or form impressions) about a candidate’s technical skills. (Marlow et al., 2013)

https://codeformore.com/how-to-write-up-open-source-experience-when-you-dont-have-any/
Knowledge gap

outider | newcomer | developer | long-term developer

motivation | onboarding | retention

This work
RESULT HIGHLIGHTS: Social capital explains sustained participation

Higher social capital

Being part of teams with more diverse information, especially for women
~40% of GitHub contributors disengage
Social capital theory for sustainable participation

Bonding social capital:
benefiting from strongly connected network

Willingness to continue
(Coleman, 1990)

Bridging social capital:
benefiting from network with diverse info

Opportunity to continue
(Burt, 1998, 2001)
H1: more social capital ~ more prolonged engagement

Bonding social capital:
benefiting from strongly connected network

Willingness to continue
(Coleman, 1990)

Bridging social capital:
benefiting from network with diverse info

Opportunity to continue
(Burt, 1998, 2001)
Cohesive networks might foster discrimination / exclusion
On GitHub, women disengage earlier than men

After one year ca. 70% of men are still active but only ca. 60% of women
On GitHub, women disengage earlier than men

After one year ca. 70% of men are still active but only ca. 60% of women

Why do contributors disengage?

- Developer survey in our paper:
  - Work-related (e.g., new job)
  - Personal* (e.g., different hobby)

*women cite more often than men

See also:
- Miller et al, OSS 2019
- Iaffaldano et al, SoHeal 2019
H2: Teams with more diverse information ~ more prolonged engagement, esp. for women

Information diversity should reduce the risk of demographic-based echo chambers.
Large-scale mixed-methods study

GitHub → Filter: 1+ commits Full name → Sample 300,000 users → Balanced sample ~30,000 F ~30,000 M → Cox regression

Small sample 1,000 users → Survey 88 replies
Inferring gender from names

https://github.com/tue-mdse/genderComputer

https://www.namsor.com

name features, e.g., the last two characters

Public name lists & celebrity names, including 3,000 East Asian names

Naive Bayes classifier

Binary gender prediction
Bonding social capital – Team Familiarity

(de Vaan et al., 2011, Lutter 2015)
Bonding social capital – Recurring Cohesion

(de Vaan et al., 2011, Lutter 2015)
Bridging social capital – Language Diversity

TIME

Java
C
C++
C++
C
Project X
End of data

(de Vaan et al., 2011, Lutter 2015)
Bridging social capital – Share of Newcomers

(de Vaan et al., 2011, Lutter 2015)
## COX regression model

<table>
<thead>
<tr>
<th>Contributor</th>
<th>Time</th>
<th>Active</th>
<th>Social capital</th>
<th>Control variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Team familiarity, Recurring cohesion, Language diversity, Share of newcomers</td>
<td>Project size, Project owner, ......</td>
</tr>
<tr>
<td>active, major</td>
<td>2008 Jan – Mar</td>
<td>True</td>
<td></td>
<td>Project size, Project owner, ......</td>
</tr>
<tr>
<td>contributor, not</td>
<td></td>
<td></td>
<td></td>
<td>Project size, Project owner, ......</td>
</tr>
<tr>
<td>owner, social</td>
<td></td>
<td></td>
<td></td>
<td>Project size, Project owner, ......</td>
</tr>
<tr>
<td>capital</td>
<td></td>
<td></td>
<td></td>
<td>Project size, Project owner, ......</td>
</tr>
<tr>
<td></td>
<td>2008 Jan – Mar</td>
<td>True</td>
<td></td>
<td>Project size, Project owner, ......</td>
</tr>
<tr>
<td></td>
<td>2009 Apr – Jun</td>
<td>False</td>
<td></td>
<td>Project Size, Not project owner, ......</td>
</tr>
</tbody>
</table>
H1: more social capital ~ more prolonged engagement
H2: Language diversity interacts with gender

Survival difference between contributors with high and low language diversity

Time in months

Gender
- Female
- Male

Difference in survival probability
What next?

Recommend projects that can help build social capital

Offer mentorship

Use badges to show a community’s culture

- Community culture: We welcome help
- Community culture: We are friendly 😊
- Community culture: <3
- % of newcomers: 30%
Social capital explains prolonged engagement

On GitHub, women disengage earlier than men
• After one year ca. 70% of men are still contributing to GitHub projects but only ca. 60% of women

Social capital explains sustained participation

Large-scale mixed-methods study

What can we do?

Recommend projects that can build social capital
Offer mentorship
Use badges to show a community’s culture

Acknowledgements:

Contact: Sophie Qiu
@sophiehsqq
hsqq@cmu.edu

@alexander_nolte
alexander.nolte@udo.edu

@aserebrenik
a.serebrenik@tue.nl

@b_vasilescu
vasilescu@cmu.edu

Code and data:

DOI 10.5281/zenodo.2566684