

Diversity-Innovation Paradox in Open-Source Software

Updates

- I will be joining Hongbo's project on measuring innovation and collaborate on future writeups.

Project backgrounds

- Diversity-Innovation Paradox in science is identified by Hofstra et al.
 - Diversity breeds innovation
 - Yet underrepresented groups have less successful careers
- Can we find similar effects of the paradox in open-source software?
- Our research questions
 - ~~How do we define and measure innovations in the context of open source software?~~
 - ~~Do contributors in more diverse groups produce more innovations?~~
 - ~~Does contributing more innovations lead to a longer open source engagement?~~
 - How does exposure to diverse groups affect individual's contribution to innovation?
 - How does producing oss innovations affect individual's prolonged participation in oss?

How to sustain open-source participation

- Backgrounds
 - Open-source software is now ubiquitous and key to the infrastructure of society
 - Vasilescu et al. shows that tenure and gender diversity is a good predictor of productivity
- Knowledge gap
 - How do groups of open-source contributors generate innovative software?
 - Does software innovation leads to sustained open-source participation?
- Objectives
 - Sustain open-source engagement of diverse groups of people
 - So that they can maintain and produce important, innovative open-source software

Study overview

1. Collect innovation for open-source projects
2. Collect diversity and calculate measures for contributors of the innovations
3. Collect open-source “career” for contributors of the innovations
4. Statistical analysis (e.g. regression) on innovation versus diversity to find potential correlation
5. Statistical analysis (e.g. regression) on innovation/diversity versus open-source career

We use World of Code to gather innovations and career

- We define software innovation as the first co-usage of two packages
 - Hofstra et al. defines scientific innovation as first co-occurrence of two concepts
- World of Code provides package usage for a vast number of open-source projects
 - Including when, where, and who introduced the package into the project
- We have collected this data only up to 2014, and plan to continue to present day
- World of Code also allows us to find participation in any projects for a given contributor

We continue to collect and calculate diversity measures

- We have gender diversity data for projects in npm registry
- But innovations also come from many projects not in the registry
 - We will collect gender diversity for these projects
- We also need to collect and calculate tenure diversity
- And we want to calculate individual metrics representing an individual's exposure to group diversity from any open-source projects they participated in

Bogdan's comments

- Sam: Cool project but the presentation felt a little unfocused. Can you narrow down the study to fewer (1?) research question but provide a sharper argument for why that question is important, and more details on how you will answer it? The connection from diversity to innovation to disengagement to sustainability was a bit hard to follow. When explaining what an “innovation” is, you could probably simplify the presentation a bit to focus on the intuition and less on the specifics (e.g., Javascript). The hackathon back story also seemed unrelated?