Computational Scientists write **great software**: **much better** than most software engineers.

Can Empirical SE be Adapted to Computational Science?

PI: Dr. Tim Menzies, North Carolina State University
http://ai4se.net/se4cs

PROBLEM

Computational Scientists study micro events within atoms that add up to predictable properties of macro materials (e.g. the sun). Similarly, empirical software engineers study micro patterns within software projects, to learn predictable properties of those projects.

Hypothesis: Integrating SE practices will help computational scientists produce better science (e.g. more reliable, more reproducible, & more efficient).

RESEARCH ROADMAP

- [x] Investigate how computational scientists conduct research
- [ ] Mining for SE issues within computational science domain
- [ ] Explore usefulness of SE in computational science
- [ ] Automating SE practices

METHODS

Literature Study

Pre-Study Community Survey

Github Mining

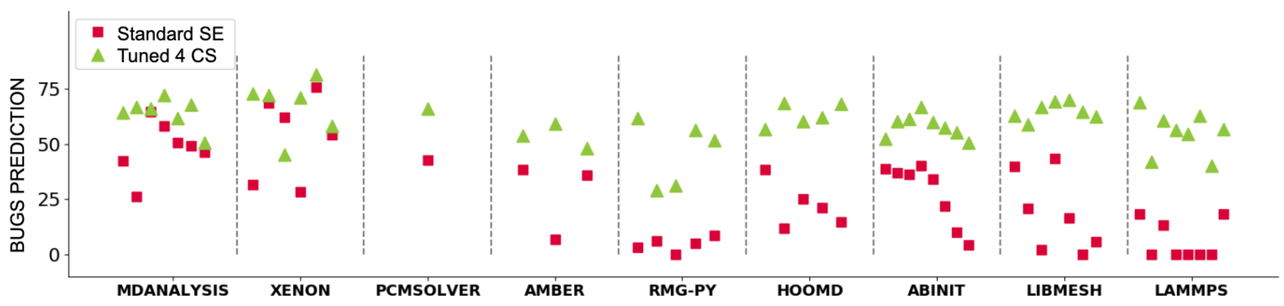
Post-Study Community Survey

WHAT CAN WE DO FOR YOU?

We seek partners to offer feedback on our quality predictors on your code.

- **Defect Forecasting:** Identify the part that are fault prone.
- **Identify Heroes/Antiheroes:** Identify effective and ineffective contributors.
- **Reviewer Selector:** Automatically identify best reviewer for a PR.
- **Test Case Prioritization:** Minimize time to failure for PRs by reordering test cases dynamically.
- **Code/API Assistant:** Jump-start new users with tailored code completion recommendations & propose bug fixes or appropriate changes while coding.

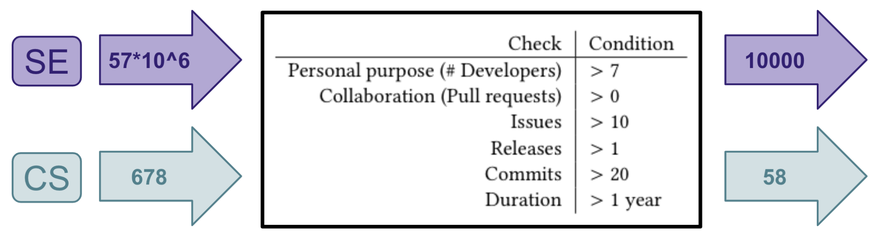
SE METHODS NEED ADAPTION FOR CS



RED SQUARE: Shows prediction results based on standard SE methods which performed badly on Computational Science project.

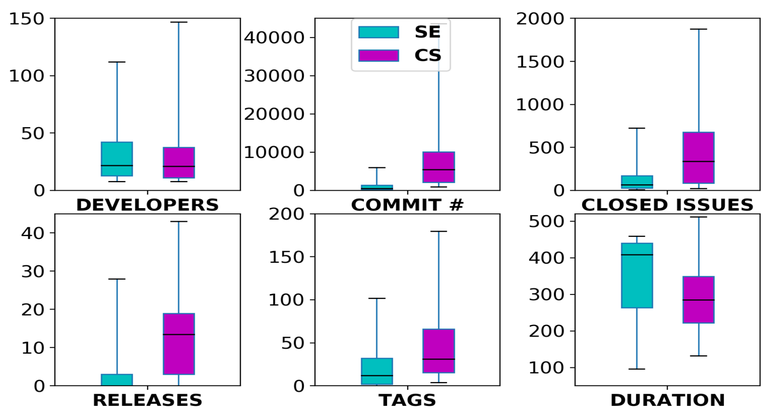
GREEN TRIANGLE: improved results after tuning Empirical SE methods.

CS DEVELOPERS: 487 TIMES MORE SERIOUS THAN SE

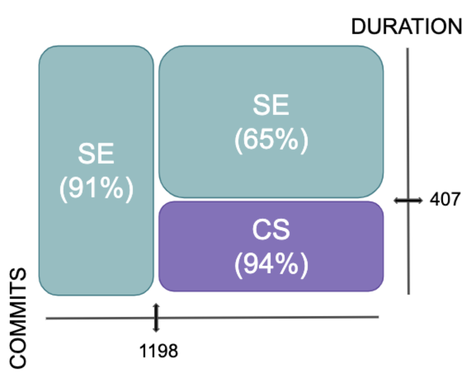


Number of CS and SE projects BEFORE and AFTER applying sanity checks.

CS CODE: MORE GRANULAR, BUILT FASTER



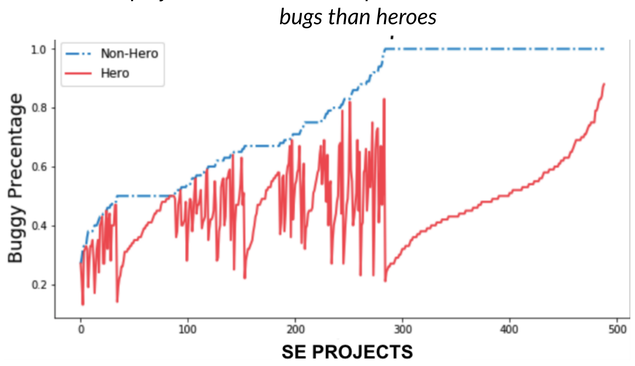
LEFT: Github statistics of 1000 SE projects & 59 CS projects



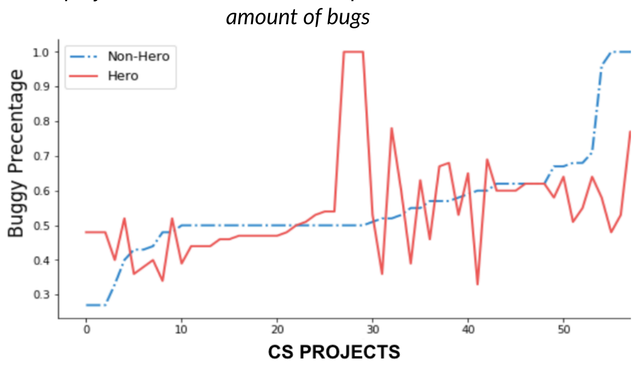
RIGHT: Categorize SE or CS projects

CS DEVELOPERS: DELIVER CODE WITH FAR FEWER BUGS

In SE projects: non-hero developers introduced 1.3-1.7 more bugs than heroes

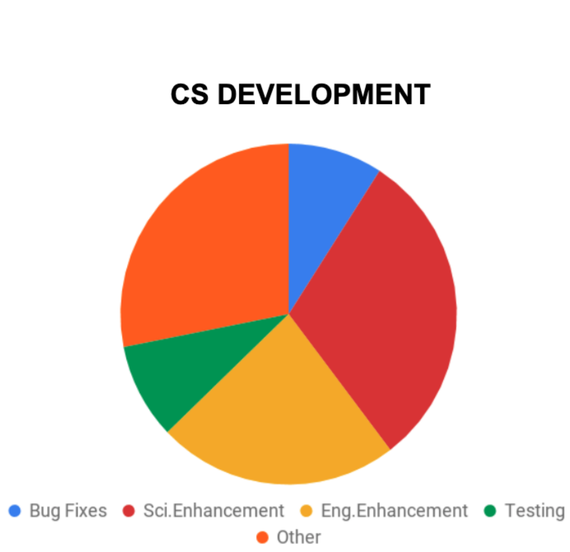


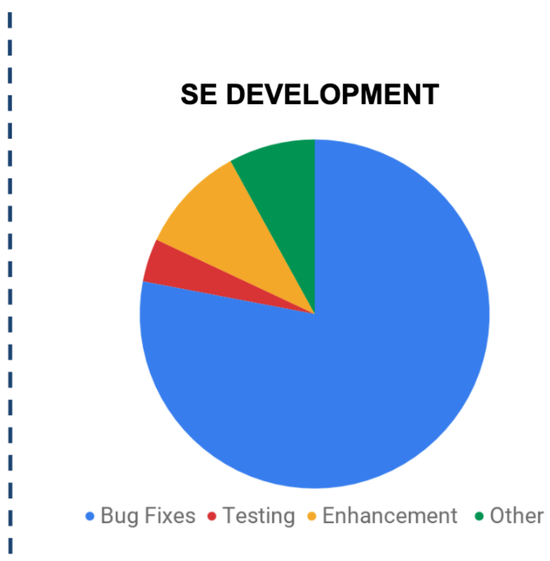
In CS projects: non-hero hero developers introduced the same amount of bugs



CS DEVELOPERS: BOLDER EXPLORERS OF NEW IDEAS

Computational scientists **focus more on new ideas**, than maintaining old ones.





Awards: #1931425

H. Tu, Z. Yu, T. Menzies, IEEE TSE, 2020
Better Data Labelling with EMBLEM
https://arxiv.org/pdf/1905.01719