Open-source community health

#project
Project SharePoint

Research questions

- Measuring open-source community health at the ecosystem scale (through metrics gleaned from repo mining)
  - pypi
  - rubygems
  - npm
  - maven
  - opam and/or haskell (if possible)
- Metrics suggested
  - **Community standards**
    - What percentage of repos have a code of conduct file?
    - What percentage have a contributing file?
  - **Responsiveness to community members**
    - Of all issues opened by outsiders, what percentage have a comment from an insider?
    - What's the average time taken for the insider's comment (time-to-response)?
  - **Community participation** (contribution dispersion; mostly concentrated among a few insiders or spread out among a large number of equally-contributing members of the community?)
    - Of all code contributions (commits or additions/deletions) by contributors, what's the spread of number of contributions? How is this distributed among insiders vs. outsiders?
    - (Gives us a sense of how much the community as a whole contributes to the package.)
  - **Welcomingness to newcomers**
    - What percentage of repos have issues marked "good for beginner" or "good first issue"
  - **Repeat contributions** — For people who made their **first** PR in a package,
    - What percentage of them made subsequent PRs within a year?
    - For those who did make more than 1 PR, on average how many PRs were made after that first one?

- **Software engineering practices** — testing, documentation, evidence of code review...? maybe even dynamic analyses for key projects (e.g., code coverage)
- **Q&A community on StackOverflow** (# questions, % with accepted answers, see literature for more measurements)

- **Diversity** (how?)