Definition of sustainability (DS)

  DS:Community: community behind this project is sustainable

  DS:Functional: package is usable by other programmers and doesn't require additional maintenance
  (DS:

  DS:Userbase: this package has a non-empty userbase

  DS:Maintained: package is able to attract contributions / has recent commits

Features: Commits (FCom)

  FCom:Depends: interpretation depends on other factors, e.g. package size and userbase

  FCom:Misleading: this measure is hardly interpretable, e.g. because of commit squashing

  FCom:Recency: only recency of commits matters

Features: Contributors (FCon)

  FCon:Positive: number of contributors has a strictly positive effect on sustainability

  FCon:Core: only size of the core team matters

  FCon:Noise: identification of noise sources

Features: Size of the Core Team (FSCT)

  FSCT:Positive: size of the core team has a strictly positive effect on sustainability

Features: Issues (FI)

  FI:Noisy: metric needs to be adjusted to submission quality

  FI:Depends: interpretation depends on other factors (size of the project/team)

  FI:Positive: the more issues reported the better

Features: Issue Reporters (FIR)

  FR:Noisy: metric needs to be adjusted to submission quality

  FR:Positive: the more reporters the better

Features: Upstreams (FU)

  FU:Depends: interpretation depends on many other factors or has a non-linear effect

  FU:Noisy: metric needs to be adjusted to quality of dependencies

  FU:Positive: upstreams are generally good for sustainability

Features: Downstreams (FD)

  FD:Positive: the more downstreams the better

  FD:Noisy: not all downstreams are helpful

  FD:Neutral: downstream projects do not give a any advantage

  FD:Negative: downstream projects require extra maintenance effort

Features: relative Position in the dependency chain (FPos)

  FPos:Positive: transitive dependencies are positive (perhaps less than direct ones)

  FPos:Neutral: transitive dependencies do not give any advantage

  FPos:Depends: effect extent is expected to depend on other factors (project class, chain length etc)

Features: Social ties (FSoc)

  FSoc:Useless: this measure is too noisy to indicate anything

Features: University Involvement (FUI)

      FUI:Positive: university projects are more sustainable as backed up by a large organization

      FUI:Noisy: effect of this metric depends on other factors (actor role, field, diversity)

      FUI:Negative: "academics write terrible code"

Features: Commercial Involvement (FCI)

      FCI:No: commercial involvement is non-existent

      FCI:Depends: commercial contributions may be both useful and harmful

Suggested Features (New)

      New:Issues: issue management practices (triaging, response timing etc)

      New:Users: existing/active userbase

      New:CI: use of Continuous Integration

      New:CR: use of Code Reviews

      New:Test: presence of autotests and related practices (e.g. cross-platform test runners)

      New:Doc: usable documentation

      New:Sup: community support (Slack, mailing list etc)

      New:Other

Part 2 - Features

I am going to ask about several features and their meaning for sustainability. Try to think in terms "If I look at this metric, what can I conclude about project sustainability?". For each feature, please explain your interpretation of this feature and how it affects sustainability. If you have extra comments or clarifications, please feel free to add.

Commits (i.e. number of commits per month)

Contributors (i.e. number of unique committers per month)

Size of the core team (i.e. number of committers doing 90% of the work)

Issues (i.e. number of reported issues per month)

Issue reporters (i.e. number of unique issue reporters per month)

Upstreams (i.e. number of technical dependencies a project has)

Downstreams (i.e. number of projects dependent on a project)

Relative position in the dependency chain (e.g. different situations similar to: "there are few projects depending on me, but there are 10K projects depending on them")

Social ties (i.e. number of other projects connected to this project through contributors)

University involvement (of any form; if you can think of several forms please explain)

Commercial involvement (e.g. if there was a magic way to measure number of commits from firms rather than individuals)

Part 3 - suggested features

Conclusion

Thank you for your time. If you are interested in the results, I can send you a copy of the paper if/when it is accepted. We are also going to share our dataset.