

A Study on Sustainability of Open Source Software Projects

山下, 一寛

<https://doi.org/10.15017/1807066>

出版情報：九州大学, 2016, 博士（情報科学）, 課程博士
バージョン：
権利関係：全文ファイル公表済



氏 名 : 山下 一寛

論文名 : A Study on Sustainability of Open Source Software Projects
(オープンソースソフトウェアプロジェクトの持続性に関する研究)

区 分 : 甲

論文内容の要旨

オープンソースソフトウェア (Open Source Software, OSS) は、そのライセンスに従って、自由に誰に対してもそのソースコードを学習、変更、頒布することのできるソフトウェアである。企業にとっては、同様の機能を持つソフトウェアを開発せずに OSS を利用することにより、ライセンス料や人件費等のコストを削減できるなどの利点がある。そのため、個人のみならず、多くの企業でも OSS を利用することが増えており、OSS プロジェクトの持続可能性は、現在、最も重要な社会的課題の一つとなっている。

持続可能性には、「プロジェクトが持続的に成長するための発展性」と「現在のプロジェクト体制が瓦解しないための組織的な安定性」の二面がある。OSS の世界では、持続可能性は、プロジェクト単体ではなく、多数のプロジェクト群から構成される社会構造の側面から評価する必要がある。ある OSS プロジェクトが生き残るか否かは、他の競合プロジェクトの存在に影響されるからである。本研究では、米国における州間人口流動調査に用いられている Magnet/Sticky という指標を用いて、OSS の持続可能性について大規模な実証的分析を行った。Magnet/Sticky では、州間の人口流動を元にどの州が魅力的で人口を増やしているのか、またどの州が停滞しているのか、をマクロ的に捉えることができる。本研究では、州を OSS プロジェクトに置き換え、OSS プロジェクト群の持続可能性を開発者人口という観点から分析した。Magnet は新規の開発者をプロジェクトに引きつける能力で発展性に当たり、Sticky は既にプロジェクトに参加している開発者を引き止める能力で安定性に当たる。本研究の貢献を以下に述べる。

第一に、OSS プロジェクトの持続可能性を定量的に評価するために、前述の Magnet/Sticky という指標を導入し、発展性と安定性の二つの観点を組合せて利用する「OSS 持続可能性評価モデル」を提案した。従来研究においても、新規開発者の獲得や既存開発者の保持の重要性については研究が行われてきたが、一つの対象プロジェクトのみに着目し、そのプロジェクト内で新規開発者がどの程度増えたか、既存の開発者がどの程度保持できているかで持続可能性を測っていた。しかしながら、OSS では開発者が自由にプロジェクトを進めているので、同様の機能を持つプロジェクトが複数、同時期に開発される場合がある。例えば、MacOSX 向けの有力なパッケージ管理ツールには、Macports や Homebrew などが存在する。このような場合、一方のプロジェクトがもう一方のプロジェクトよりも極端に人気がある場合、人気のない方のプロジェクトは十分な開発者を得ることが出来ず、中断せざるを得ない可能性がある。そのため、プロジェクト単体で考えるのではなく、プロジェクト間の関係を考慮することは重要である。そこで本研究では、人口流動調査の分野で用いられている Magnet/Sticky という指標を、初めて OSS の研究分野に導入し、プロジェクト間の関係性を考慮した持続可能性の評価を可能にした。ただ、米国における州人口と OSS 開発者人口は定義の仕方が異なるため、Magnet/Sticky の定義をそのまま OSS プロジェクトに適用することは

できない。州人口の場合はある人は一人としかカウントされないが、OSS 開発者人口の場合は複数のプロジェクトに同時に参加することができるので、一人が複数回カウントされる可能性がある。このような OSS の特徴を考慮し、Magnet は全ての新規開発者のうち対象プロジェクトが獲得した新規開発者の割合、Sticky はプロジェクトに引き続き貢献し続けた開発者の割合、と定義し OSS プロジェクトの持続可能性分析に適用した。

第二に、OSS 持続可能性評価モデルの有用性を、大規模なデータセットを用いて実証的に示した。本研究では、発展性として Magnet を、安定性として Sticky を計測し、代表的な大規模 OSS リポジトリである GitHub 上で開発されている 1 万 5 千件以上のプロジェクトのデータを分析した。その結果、Magnet と Sticky の双方が高い場合、OSS プロジェクトが持続する可能性は非常に高く、双方が低い場合、高い確率で中断することが明らかになった。例えば、Linux Kernel プロジェクトや Mac OSX 向けのパッケージ管理ツールの Homebrew などが Magnet と Sticky が共に高かった。一方で、CSS フレームワークである Blueprint は Magnet と Sticky が共に低く、実際に 2011 年以降開発が止まっている。また、新しい開発者は一部の限られた OSS プロジェクトにその多くを獲得されており、その分布は非常に偏っているという現象が明らかになった。新しい開発者を多く獲得している OSS プロジェクトはその知名度が高いのみならず、新しい開発者が参加しやすいような開発上の工夫が行われている場合があることもわかった。一方で、持続的に貢献している開発者が非常に多いプロジェクトは、研究機関のプロジェクトである場合や、そのプロジェクトを利用している企業の開発者が貢献している場合などが見受けられた。このように、OSS 持続可能性評価モデルの導入により、OSS プロジェクトの成長の様子や、特徴的なイベントの影響などを定量的に評価できることが明らかになった。

第三に、プロジェクト組織の安定性と関係が深いと考えられるコア開発者 (Core Developer) 比率について大規模な実証分析を行った。コア開発者というのは OSS プロジェクトの中で中心的な役割を果たしている開発者であり、ある研究ではコア開発者が全体の 8 割のコードを書いていると報告している。そのため、持続して貢献を続けている開発者の中でもコア開発者は特に重要であると考えられる。従来研究では、成功し長期にわたって持続しているプロジェクトのコア開発者の割合は開発者全体の 2 割程度 (Pareto の法則に従う) という成果を報告している。しかし、これらの研究の問題点の一つとして、その成果が少数のプロジェクトをベースとした事例研究だけに依っており、一般的に Pareto の法則が成り立つかは示せていない。そこで、本研究では GitHub で開発が行われている 2,000 件以上のプロジェクトを用いて大規模な分析を行った。その結果、従来研究とは異なり Pareto の法則が成り立たないプロジェクトが多く存在し、コア開発者の割合はその開発期間やサイズに依らず非常に小さい割合から大きい割合の場合もあることが明らかになった。しかし、割合ではなく、人数で考えた場合は 5 人以下であることがほとんどであった。

本研究の結果、OSS 持続可能性評価モデルを用いて、持続的に発展している OSS プロジェクトとそうでないプロジェクトを上手く特徴づけられることが明らかになった。また、Magnet/Sticky の指標を用いることにより、導入を検討している OSS プロジェクトの安定性と発展性を定量的に測定することが可能となった。