

Corporate Dominance in Open Source Ecosystems: A Case Study of OpenStack

Yuxia Zhang*
Beijing Institute of Technology
Beijing, China
yuxiazh@bit.edu.cn

Klaas-Jan Stol
Lero, the SFI Research Centre for Software
University College Cork
Cork, Ireland
k.stol@ucc.ie

Hui Liu*
Beijing Institute of Technology
Beijing, China
liuhui08@bit.edu.cn

Minghui Zhou
School of Computer Science, Peking University
Key Laboratory of High Confidence Software
Technologies, Ministry of Education
Beijing, China
zhmh@pku.edu.cn

ESEC/FSE 2022 Research Papers “Corporate Dominance in Open Source Ecosystems: A Case Study of OpenStack”

発表者：田中 智達

背景

- OSSは情報技術や社会の基盤で、数十億ドル規模のビジネスとも言われている[72].
- ソフトウェアスタック全体の付加価値部分にリソースを投資しなければならないという認識を強めている [21, 68].
 - そのためOSSコミュニティやエコシステムに参加する企業数は増加しており、中には OSSコミュニティの主要な貢献者になる企業もある[26, 58, 78].
 - それらの企業は常にビジネス目標によって動かされている [24, 42][78, 84].

企業によるOSS関与

- このような企業の強い関与は、OSSプロジェクトの性質やコミュニティを変えてしまう可能性が高い。
 - ボランティアや他の貢献する企業を締め出す可能性.
 - 従来の強みであるオープンな開発モデルが消失する可能性.
- プロジェクトの失敗や放棄の割合が非常に高いことを示唆する研究もある[8, 57, 72].
 - 企業の優位性とOSSプロジェクトの持続可能性の関係をより良く理解することが重要.

関連研究

- 貢献する企業が他の企業(企業とボランティアの両方)と協力することが前提.
 - 企業の支配や, 実際にOSSのコミュニティへの影響を調査した研究は非常に少ない.
- 多くは以下のことに焦点が当てられている.
 - 企業がなぜOSSプロジェクトに参加するのか[71].
 - 企業がOSS製品やサービスを通じてどのように利益を上げるのか[27].
 - 企業がどのようにOSSプロジェクトに貢献し影響を与えるのか[77, 79, 84].

先行研究

- 利益相反によるOSSコミュニティの緊張の証拠の提供
 - オープンソース開発者の離脱や、企業のプロジェクト放棄につながる可能性を示唆[54, 75].
- OSSプロジェクトの持続可能性についての調査[3, 46, 53, 67]
 - 企業参加によるOSSプロジェクトの持続性への影響はほとんど考慮されていない[67].
- OSSのエコシステムにおける企業支配の実態と、それがプロジェクトの持続可能性に及ぼす影響を理解することは、支配の動向をよりよく理解し、適切な行動をとるために非常に重要.

RQ

- RQ1

- OpenStackにおける企業優位性はどの程度浸透しているのか.

- RQ2

- OpenStackにおける企業の優位性のパターンとは何か.

- RQ3

- 企業の優位性はOSSリポジトリの存続と相関があるか.

OpenStack

- クラウドコンピューティングのインフラストラクチャのオープンソース・フリーソフトウェア.
- 本研究で選出した理由として以下が挙げられる.
 - 2010年7月に設立され, 非常に広範な開発の歴史があり, 有意義な分析を可能にする十分なコミットのメタデータが利用可能である.
 - OpenStackのエコシステム内の1000以上のレポジトリの開発に何百もの企業が参加している.
 - 支配パターンを研究するための豊富なデータと詳細な生存分析のための十分なデータを提供.
 - OpenStackはいくつかの先行する科学的調査の対象である.

データセット

- 先行研究によって提供されたデータセットを使用[78, 79].
 - 2010年7月から2019年1月までの開発履歴をカバーする OpenStackの加工データセット.
- 開発者とコミットの所属の正確性については、該当する開発者に直接アンケートを実施することで検証、高い精度(93%)でアーカイブされている[78]ため高いと言える.
- このデータセットでは490 社が1,216件のOpenStack リポジトリに貢献し、18の完全なバージョンを進化させ、その結果338,035のコミットを記録.

企業支配力の測定

- 経済学の古典的な理論である市場支配力理論に基づいている[29].
 - 特定の市場において、競合する製品に比べ大きな力の部分を支配している企業、ブランド、製品、サービスに関する理論 [2, 29]であり、最も一般的に利用される尺度である.
- 市場における支配力の評価は通常、Melnikらによる市場支配力指標によって特徴付けられ[32, 39], 次のように計算される.

$$s^D = \frac{1}{2} \left[1 - \gamma (s_1^2 - s_2^2) \right]$$

s_1, s_2 : 上位の市場占有率($s_1 > s_2$).

γ : 潜在的な競争の産業別・時間依存的な重要性を捉えるもの.

s^D : 上位1社が市場で支配的地位を有するとみなされる閾値の市場シェア.

企業支配力の測定

- 支配力指標は、主に市場における2大企業の市場シェアに着目。
 - 最も近い競合企業の規模に依存した尺度。
- この尺度は適用が容易であるため、広く使用することができる[32].

$$s^D = \frac{1}{2} \left[1 - \gamma (s_1^2 - s_2^2) \right]$$

s_1, s_2 : 上位の市場占有率($s_1 > s_2$).

γ : 潜在的な競争の産業別・時間依存的な重要性を捉えるもの.

s^D : 上位1社が市場で支配的地位を有するとみなされる閾値の市場シェア.

企業支配力の測定

- 市場シェアをコードオーサーシップに置き換え，企業支配力を算出することを提案.
 - モジュールの保守に責任を持つ開発者を特定するための多くの研究で広く利用されている[5, 41, 66].
- あるバージョン v におけるOSSリポジトリ r の企業 c のCA (Code Authorship)は，以下のように計算される.

$$CA(c, r, v) = \frac{\text{バージョン}v\text{の開発で}c\text{社がリポジトリ}r\text{に貢献したコミット数}}{\text{バージョン}v\text{に含まれるリポジトリ}r\text{に貢献したコミット数}}$$

- 各企業のコードオーサーシップを決定したのち，先ほどの指標に代入することで企業の支配力を評価.

$$CA^D(r, v) = \frac{1}{2} \left[1 - \left(CA_{1,r,v}^2 - CA_{2,r,v}^2 \right) \right]$$

企業支配力の測定

- $CA_{1,r,v}^2$ はバージョン v におけるリポジトリ r への貢献度が最も高い企業のコードオーナーシップを表し, $CA_{2,r,v}^2$ は次点の企業のコードオーナーシップを表す.
- $CA_{1,r,v}^2$ と $CA^D(r, v)$ を比較し, 企業優位性の有無を判定.

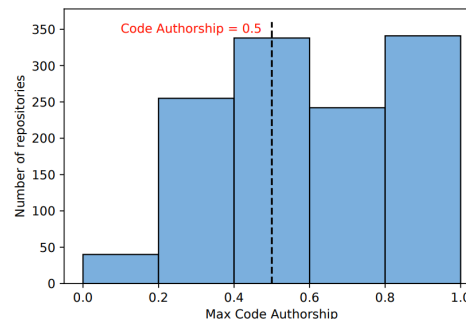
$$CA^D(r, v) = \frac{1}{2} \left[1 - \left(CA_{1,r,v}^2 - CA_{2,r,v}^2 \right) \right]$$

$CA_{1,r,v}^2 > CA^D(r, v)$	首位の企業はバージョン v の開発において、リポジトリ r では支配的な地位を有している.
$CA_{1,r,v}^2 < CA^D(r, v)$	このリポジトリを支配している企業はなく、無関係な個人からなる組織により支配.

- 無関係な個人からなる組織というのはボランティア主導のOSSコミュニティという「伝統的」な概念にあたる.

RQ1

- 定義された公式に従い，すべての貢献企業のコードオーサiershipを計算.
 - 1社で占められているリポジトリが888 件(73%)あり，コードオーサーシングの中央値は70.6%であった.
- OpenStackのリポジトリ開発では，全開発期間中，また特定のバージョンのOpenStackにおいても，企業による支配が一般的.
- 70%以上のリポジトリが4分の1の企業によって支配され，そのうちの半数は複数のリポジトリを支配.

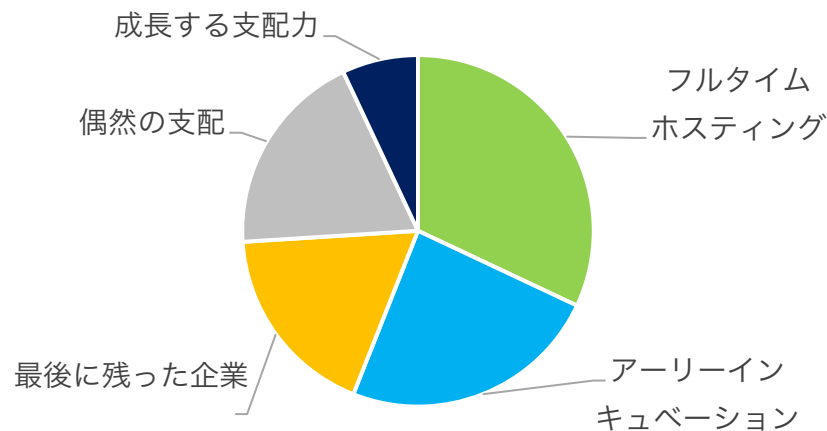


図：コードオーサiershipの分布

RQ2

- ランダムに選んだ120のリポジトリの開発履歴を手動で分析.
 - 120のリポジトリの開発履歴は、2バージョンから18バージョンまで様々であり、中央値は8バージョン.
- 5つの企業支配のパターンを特定.
 - いくつかの支配パターンは、リポジトリの種類と相関がある.

図：5つの支配パターンの分布図



RQ3

- 各レポートリの生存バージョンについて、支配状態その他の制御変数の値を計算.
- 675のイベント(休眠状態のレポートリ)を含む、6,808の観測値を獲得. 評価はCox比例ハザード回帰モデルを使用.
- 既存の要因(コミット数, レポートリカテゴリ)考慮しても、企業の優位性とレポートリの生存確率の間に統計的に有意な関係があることが判明.
 - 1社に支配されることは、OSSレポートリの長期的な発展に有害である可能性がある.
- OSSへの企業の参加について、より慎重な監視と管理が必要.

結論

- OSSエコシステムへの企業の参加が進むにつれ、エコシステムの将来の進化に対する企業の影響力も強くなる.
- その結果エコシステムの持続可能性に大きな影響を与える可能性がある.
- 本研究ではOpenStackを対象に以下のことを示した.
 - 企業の優位性が一般的な問題である.
 - 企業の優位性には5パターンなり、それを特定.
 - リポジトリの大きなセットに対して生存分析を適用し、リポジトリの企業の優位性がその生存確率に負の相関がある.
- 今後、OSSコミュニティがこの問題と関連するリスクに対する認識を高め、将来の持続可能性を確保するためにガバナンスメカニズムを調整するのに役立つと思われる.