

An Analysis of Funding and Its Impact on Development Activities in Open Source Projects

Gakushi Kimura · Olivier Nourry · Yoshiki Higo (UOsaka)



研究背景

OSS(オープンソースソフトウェア)は世界中の企業、行政、開発者に利用されている

- Linux
- OpenSSL
- Log4j
- Python

開発が停止すると**世界中のサービスに影響が出る**

OSS維持の難しさ

1. 多くのOSSは少数の開発者に依存^[1]
2. 脆弱性が見つかってもすぐに対応できない可能性がある
3. 脆弱性が悪用される可能性がある
Heartbleed^[2]やLog4Shell^[3]など

[1] O Nourry, M Kondo, S Saito, Y Iimura, N Ubayashi, Y Kamei.

Myth: The loss of core developers is a critical issue for OSS communities

[2] Marquess, S. (2014, Apr, 12). Of Money, Responsibility, and Pride, Speeds and Feeds.

<http://veridicalsystems.com/blog/of-money-responsibility-and-pride/index.htm>

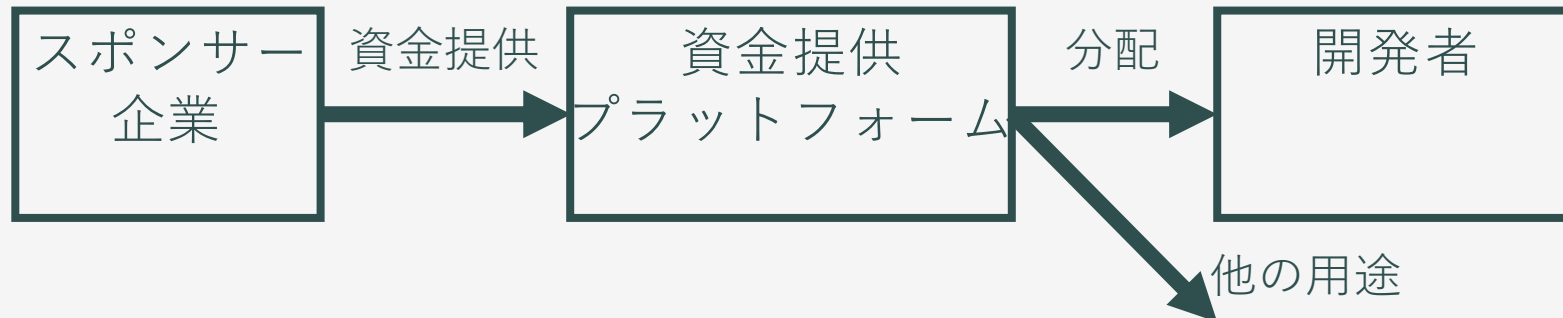
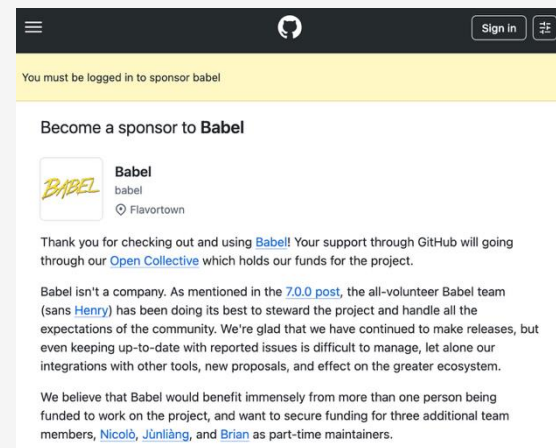
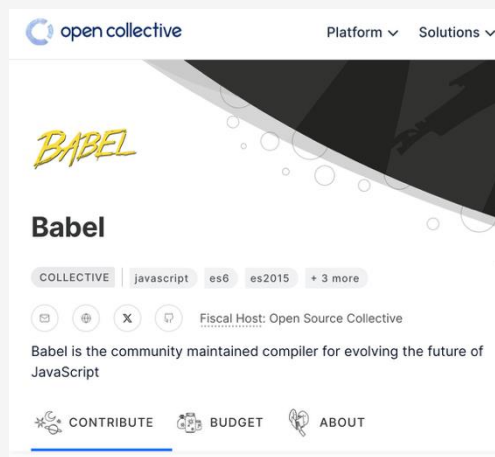
[3] MIT TECHNOLOGY REVIEW, "Log4J問題で明るみになった「ボランティア頼み」の危うさ,"

<https://www.technologyreview.jp/s/263829/the-internet-runs-on-free-open-source-software-who-pays-to-fix-it/>

OSSへの資金提供

開発者へ寄付をする企業、団体が増加し
資金提供プラットフォームが登場した

- Open Collective
本研究ではこちらを使用
- Github Sponsors



本研究の目的

OSSへの資金提供の実態を明らかにし、
Open Collectiveへの参加と資金の使い方が
開発・保守活動とどのように関連するかを調査する

研究対象

- Open Collective上の取引履歴
- GitHub上のコミット/Issue/PR(Pull Request)履歴

研究方法

1. Open Collectiveの取引データをマイニング
2. 資金の流れを集計
3. LLMを使用して資金の使途を分類
4. Open Collectiveに登録されているプロジェクトと対応するGitHubリポジトリをクローン
5. 各プロジェクトの資金の使途とコミット数、Issue/PR数を取得
6. 資金提供前後で比較

マイニングしたデータについて(SQLテーブル)

- Open Collectiveに登録されているOSSのデータ

id	プロジェクト名	Webサイト	Github リポジトリ
⋮	⋮	⋮	⋮

- Open Collectiveの取引データ

説明文	時刻	金額	送金元 アカウント	送金先 アカウント	種類 (寄付/支出)
⋮	⋮	⋮	⋮	⋮	⋮

- GitHubのコミット、Issue、PR履歴

リポジトリ名	時刻	種類 (コミット/Issue/PR)	メッセージ	ラベル	アカウント
⋮	⋮	⋮	⋮	⋮	⋮

取引データについて

- 資金を受け取った記録(寄付)
 - 資金を使った記録(支出)
- の2種類ある



リサーチクエスチョン

RQ1: OSSへの資金提供の規模はどの程度か

RQ2: 資金は誰から誰へ流れているか

RQ3: Open Collectiveへの参加と
全支出に占める開発関連支出の割合は
開発、保守活動とどのように関連しているか

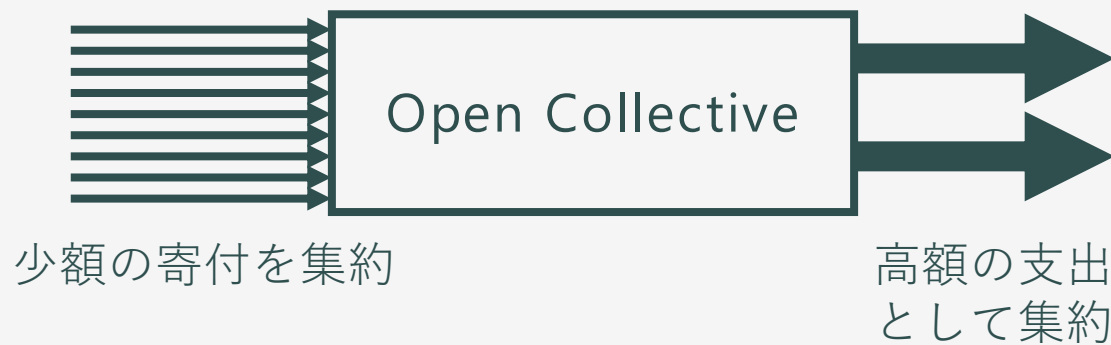
RQ1: OSSへの資金提供の規模

寄付と支出について
それぞれの資金の規模を調査した

- 寄付: 少額、高頻度
 - 支出: 高額、低頻度
- であることがわかる

寄付と支出それぞれの規模

指標	寄付	支出
回数	108万回	5万回
合計金額(USD)	6,185万	7,911万
金額の中央値(USD)	5	200



RQ1: 支出の用途

プロジェクトの支出データの説明文をLLMに与え
用途をラベル付けした

支出データの分類結果

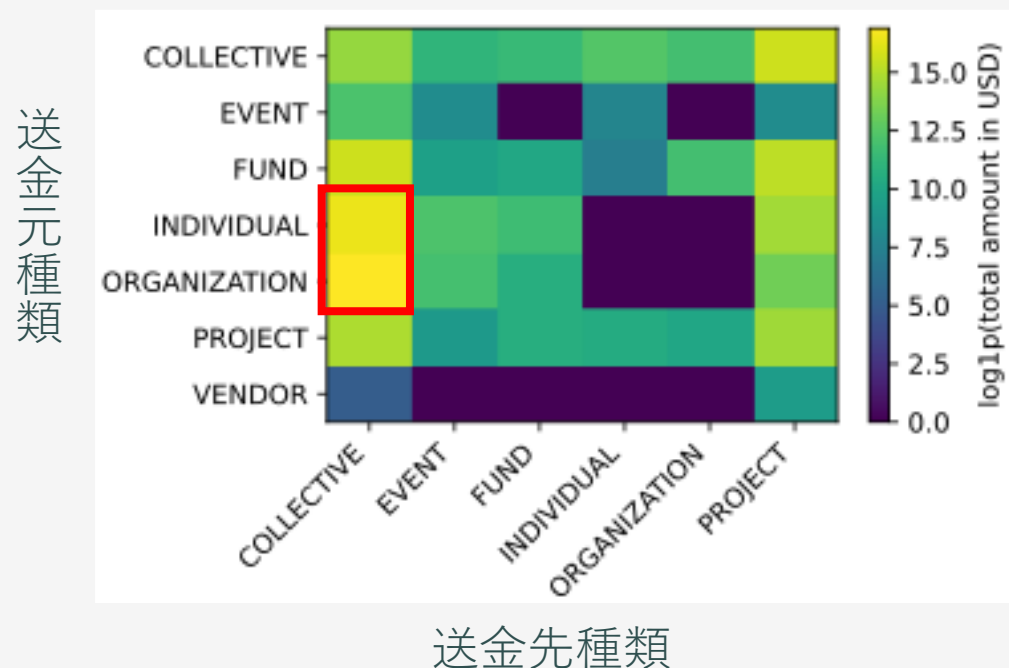
ラベル	支出の回数	割合(金額ベース)
開発	920	44.5
インフラ、月額サービス	630	2.7
マーケティング、イベント	221	3.8
開発以外の活動	131	8.3
備品	85	1.2
食べ物、消耗品	84	0.9
不明	92	38.6

開発に最も多くの金額を使用していることがわかる

RQ2: 資金は誰から誰へ流れるか(寄付)

取引データをアカウント種類毎に集計
個人や組織からプロジェクトへ流れている

- INDIVIDUAL
個人
- ORGANIZATION
組織、スポンサー企業
- COLLECTIVE
複数のプロジェクトが
合わさったもの

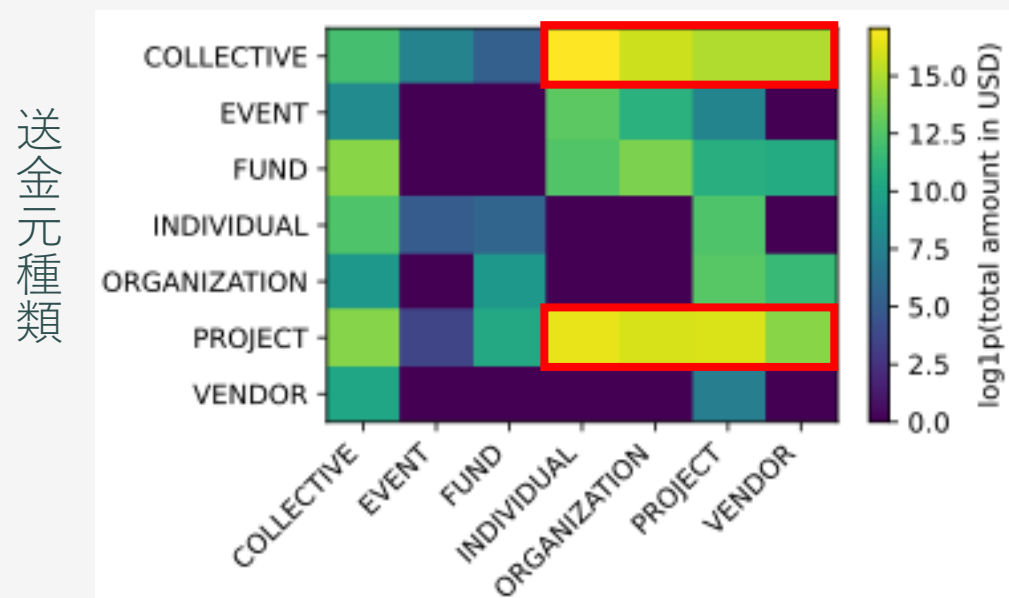


資金の流れのヒートマップ(寄付の場合)

RQ2: 資金は誰から誰へ流れるか(支出)

プロジェクトから個人、別プロジェクト、ベンダーに流れる

- INDIVIDUAL
個人
- ORGANIZATION
組織、スポンサー企業
- COLLECTIVE
複数のプロジェクトが
合わさったもの
- VENDOR
SaaS、ECサイトなど

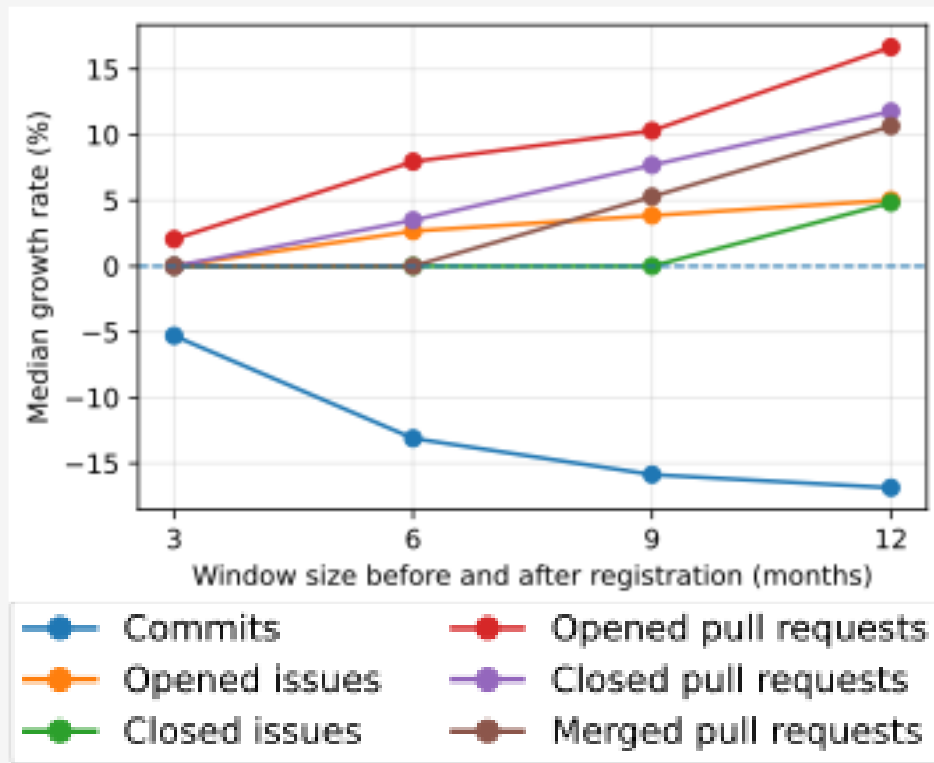


資金の流れのヒートマップ(支出の場合)

RQ3: 登録後Issue/PRは増加、コミットは減少

Open Collective登録前後の各指標の増加率の中央値をまとめた

Issue/PRは有意に増加
一方コミット数は減少



Issue、PR、コミット数の増加率の中央値

RQ3: 開発関連支出の割合とコミット増加率

各プロジェクトについて、支出のうち
開発関連支出の割合を計算

割合の上位33%、中位33%、下位33%で3群に分ける
それぞれの群について、Open Collective加入前後の
コミット数の増加率を計算

支出に占める開発関連支出の割合とコミット数増加率の関係

	開発関連支出額(平均)	コミット数増加率(中央値)
上位33%	18,983USD	+6.1%
中位33%	82USD	-3.6%
下位33%	0USD	-18.2%



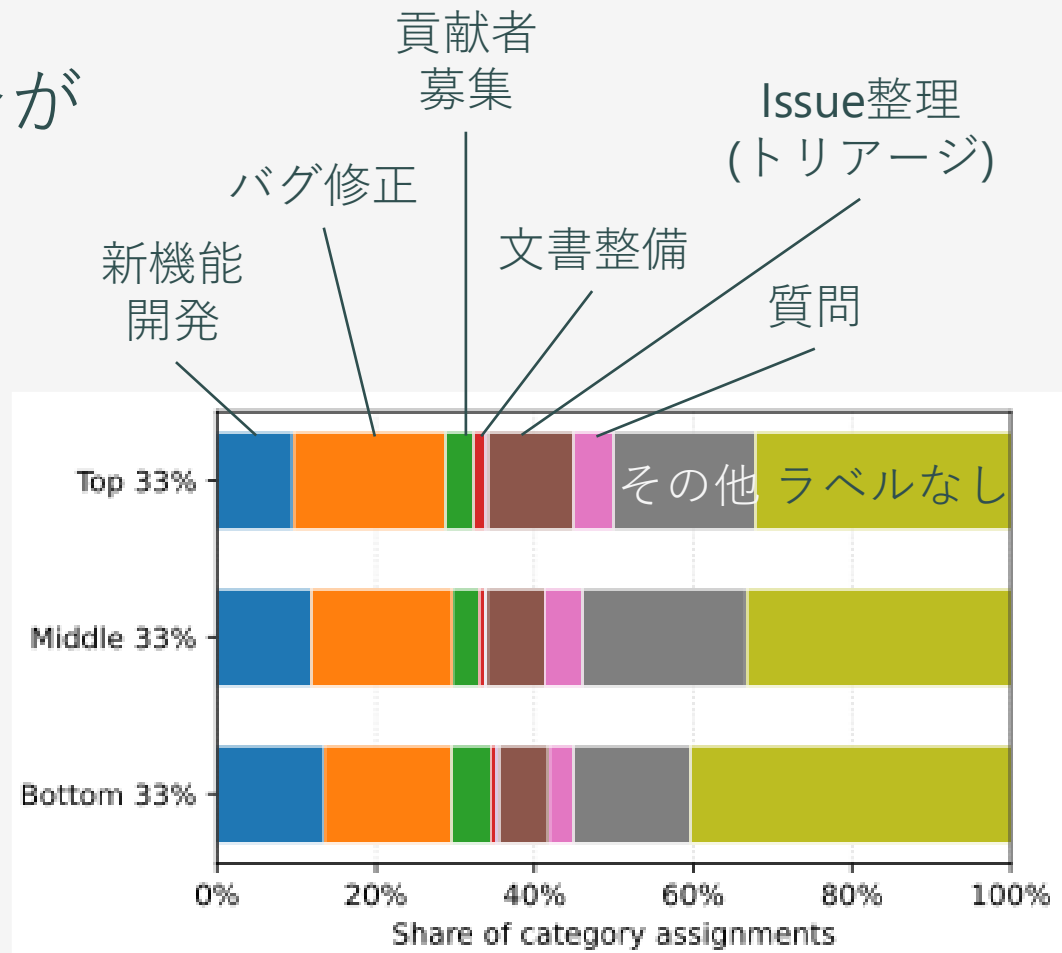
有意に
増加

RQ3: 開発関連支出の割合とIssueラベル

開発関連支出の割合が高い群では

- バグ修正
- 文書整備
- Issue整理
- 質問

の割合が増加した
(有意差なし)



Issueラベルの分類の三分位比較

考察

- OSS支援では資金を集めることだけでなく「**何に使われるか**」が重要
- **開発関連の支出割合が高い**プロジェクトでは、開発、メンテナンス活動が活発になりやすい可能性
- ただし、**全体的なコミット数は減少**傾向にある

まとめ

Open Collectiveの取引データとGitHubデータを用いて、OSSへの資金提供と開発活動を分析した

RQ1: 多数の少額寄付が集約され、少数の高額支出として**再分配**されていた
支出のうち、**開発関連**の支出が最も多かった

RQ2: 資金は組織、個人から集まり、支出は個人へ向かっていた

RQ3: **全体的な傾向: 登録後Issue/PRは増加し
コミットは減少
開発関連支出の割合が多い群では
コミットが増加する**