

How Do Companies Collaborate in Open Source Ecosystems? An Empirical Study of OpenStack

Yuxia Zhang , Minghui Zhou, Klaas-Jan Stol
Jianyu Wu , Zhi Jin

(Department of Computer Science and Technology,
Peking University Key Laboratory of High Confidence
Software Technologies, Ministry of Education)

ICSE '20: Proceedings of the ACM/IEEE 42nd International
Conference on Software Engineering Pages 1196–1208

発表者：西山 知弥

背景

- OSSエコシステムの発展には, 企業間のコラボレーションが非常に重要
- OSSエコシステム内での企業のコラボレーションに焦点を当てた研究は少ない
- 密接な協力関係にある企業もあれば, ほとんど交流のない企業もある[52, 86]. この現象の理由はまだ研究されていない

[52] Johan Linåker, Patrick Rempel, Björn Regnell, and Patrick Mäder. 2016. How firms adapt and interact in open source ecosystems: analyzing stakeholder influence and collaboration patterns. In International Working Conference on Requirements Engineering: Foundation for Software Quality. Springer, 63–81.

[86] José Apolinário Teixeira, Salman Qayyum Mian, and Ulla Hytti. 2016. Cooperation among competitors in the open-source arena: The case of OpenStack. In International Conference on Information Systems (ICIS). Association For Information System.

目的

- 大規模なOSS エコシステムにおいて, 企業がどのようにコラボレーションしているかを理解すること

アプローチ

- OpenStackのバージョン管理データを使って,企業の共同作業を定量化
- クラスター分析を用いて,企業間の協力関係や企業のプロジェクトへの貢献度を表すネットワークを作成

Open Stack

- クラウドコンピューティング・プラットフォームを構築・管理するためのOSSプロジェクトの集合体
- 2,000万行以上のコードで構成され, 194カ国の10万人以上の貢献者, 600社以上の企業からサポートを受けている(2019年7月時点)
- 活発で成熟したエコシステムであり, 約10年間にわたって活発に開発が行われてきたため, 非常に長いコミット履歴が確保

リサーチクエスチョン

RQ1 :

企業はどのようにしてOpenStackのプロジェクトに参加しているのか

RQ2 :

OpenStackのエコシステムには、どのようなコラボレーションのパターンが存在するのか

RQ3 :

エコシステムにおける企業のコラボレーションは、企業の生産性を向上させるのか

データセット

- GitHub APIを用いて,
2010年7月21日～2019年1月16日 のコミットメタデータを
取得
- ボット・アカウントを収集し, これらのアカウントから
送信されたコミットを削除
- 同じ開発者に属する複数の ID をマージする

Table 1: Statistics of dataset before and after cleaning

	# Projects	# Commits	# Developers	# Companies
Before	1,292	383,664	13,836	n/a
After	1,292	338,035	9,653	602

データセット

- 開発者の所属先と「開始日」,「終了日」をOpenStackコミュニティメンバーのリスト[33]を使用して特定
- 残りの1割は所属が確認できない開発者については,メールアドレスを目視

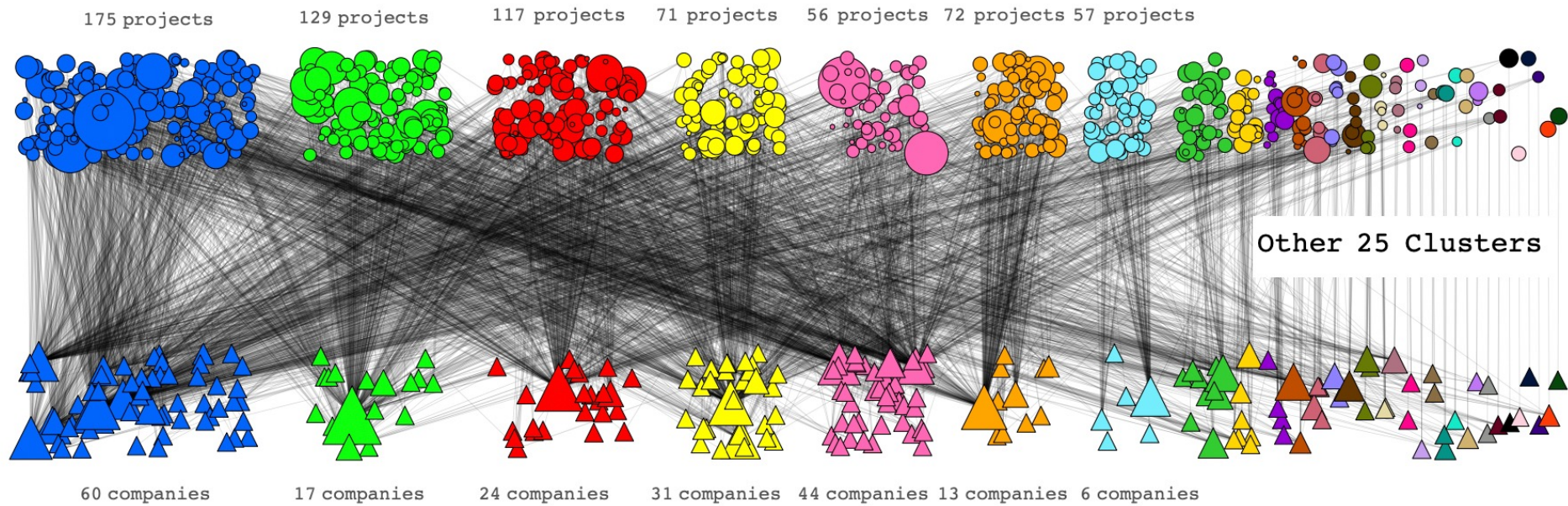
ex) redhat.com → 企業

gmail.com → ボランティア

Table 2: Example of a developer's affiliations

ID	Identities	Affiliation	Start Date	End Date
1	Monty Taylor	Red Hat	2016-06-13	2019-01-16
		IBM	2015-08-17	2016-06-12
		HPE	2011-11-21	2015-08-01
		Rackspace	2010-07-06	2011-11-20

クラスター



- 企業とその企業が参加しているプロジェクトの組み合わせを取得
- 2つの企業とその企業が参加しているプロジェクトの関係を分析することで、いくつかのコラボレーションパターンを特定

結果:RQ1

Table 3: Four business strategies of company participation in OpenStack projects

Business strategy	Description	# Companies	Example
Full Solutions (FS)	Providing full cloud solutions to users, including private/ public/ hybrid cloud services, deployment, and maintenance services, etc.	15	Rackspace
Partial Solutions (PS)	Providing solutions to users only on the basis of one or two project(s) in OpenStack.	3	Tesora
Business Integration (BI)	Integrating OpenStack with their own business	12	Intel
Complementary Services (CS)	Providing complementary services, e.g., consulting and training services around OpenStack.	2	CCIN2P3

- 企業がOpenStackのプロジェクトに参加する際,いくつかの典型的な組み合わせが発生
- BI戦略をとる企業は, 自社のビジネスとOpenStackを統合するためのプラグインやドライバーに貢献することが多い
- FSで利益を上げている企業は、デプロイメントツールに貢献することが多い
- 大手IT企業を中心としたクラスターでは, 企業規模が大きく, プロジェクトも多岐にわたる

結果 : RQ2

Project category	Description	# Projects	Example
Computing	To implement services and associated libraries regarding computing.	17	Nova
Networking	To provide capabilities for managing dynamic host configuration protocol (DHCP), static Internet protocols, or virtual area networks.	73	Neutron
Deployment	To deploy OpenStack production environments, and make it scalable to operate and upgrade.	328	Kolla
Orchestration	To provide interface and tools for the management of OpenStack services.	52	Heat
Billing	To provide billing solution by collecting data generated using cloud services and applying rating rules to calculate prices.	7	Cloudkitty
Documents	To document guides which can help users to install and use, and help contributors to participate; to document the requirements collected from all the OpenStack users.	22	openstack-manuals
Architecture Optimization	To optimize the code architecture by developing and sharing common libraries.	39	oslo
Data Analytic	To implement services and libraries about database, data processing and searching, to realize workload balancing.	20	Trove
Optimization/Policy Tools	To provide services for performance analysis and optimization of OpenStack.	19	Congress
Plugin/Driver	To facilitate integration of OpenStack and other hardware or software infrastructure.	15	Starlingx

結果: RQ2

- 競合他社であっても他者とのコラボレーションを行っている
- 積極的かつ意図的に行われるコラボレーションもあれば, 受動的かつ偶発的なコラボレーションもある
- 企業によっては, 特にその企業の基本的な製品や独自の関心事に特化したプロジェクトでは, 単独で貢献している場合もある

結果:RQ3

- 企業のコラボレーションの度合いと企業の生産性との間に, 正の相関がある

*決定的な理由はまだ不明→今後の課題

今後の課題

- 企業は **OSS** エコシステム内での自分の役割と立場を再確認し, コラボレーションパターンを利用して, 自分のコラボレーションネットワークと戦略を評価し, さらに発展させることができる

まとめ

- 多くのプロジェクトで構成されるOSSエコシステムにおけるコラボレーションのパターンを研究した研究は少ない
- このギャップを解消するために, OpenStackのエコシステムを対象とした研究を行った
- OSSエコシステムへの企業の参加が業界の常識になりつつある中, 本研究の結果は企業のOSS戦略の一助となる可能性

所感

- 他のOSSのエコシステムについても当てはまるのか気になった
- コラボレーションの傾向を知っておくと, OSSが入っているのを検出する時に役立ちそう