

How do companies collaborate in open source ecosystems? An empirical study of OpenStack

オープンソースエコシステム内での 企業間のコラボレーション OpenStackの実証的研究

著者

Yuxia Zhang, Minghui Zhou, Jianyu Wu, Zhi Jin (Peking University),
Klaas-Jan Stol (University College Cork)

1. 背景

- Linux, Android, OpenStackなどの有名なオープンソースエコシステムは複数の企業が開発に貢献^[45]
- 特定の専門領域を持つ様々な企業の貢献はオープンソースエコシステムの効率的な開発に大きな利益^[53, 80]
- 企業がプロジェクトへの貢献を行うかどうかはオープンソースエコシステムの持続可能性に大きな影響^[47, 54, 98]

したがって、企業間のコラボレーションは
オープンソースエコシステムの開発にとって非常に重要

[45] Corbet Jonathan and Kroah-Hartman Greg. 2017. 2017 Linux Kernel Development Report. <https://www.linuxfoundation.org/2017-linux-kernel-report-landingpage/>.

[47] Nicolas Jullien, Klaas-Jan Stol, and James Herbsleb. 2019. A Preliminary Theory for Open Source Ecosystem Micro-economics. In Towards Engineering Free/Libre Open Source Software (FLOSS) Ecosystems for Impact and Sustainability. Springer, 49–68.

[53] Sten R Ludvigsen, Andreas Lund, Ingvoll Rasmussen, and Roger Säljö. 2010. Learning across sites: New tools, infrastructures and practices. Routledge.

[54] Björn Lundell, Jonas Gamalielsson, Stefan Tengblad, Bahram HooshyarYousefi, Thomas Fischer, Gert Johansson, Bengt Rodung, Anders Mattsson, Johan Oppmark, Tomas Gustavsson, et al. 2017. Addressing Lock-in, Interoperability, and Long-Term Maintenance Challenges Through Open Source: How Can Companies Strategically Use Open Source?. In IFIP International Conference on Open Source Systems. Springer, Cham, 80–88

[80] Terje Snarby. 2013. Collaboration Patterns among Commercial Firms in Community-Based OSS Projects. Master's thesis. Institutt for datateknikk og informasjonsvitenskap.

[98] Minghui Zhou, Audris Mockus, Xiujuan Ma, Lu Zhang, and Hong Mei. 2016. Inflow and retention in oss communities with commercial involvement: A case study of three hybrid projects. ACM Trans Softw Engineer Methodol 25, 2 (2016).

1. 背景 | 先行研究

- コラボレーションに関する先行研究^[17, 18, 78, 79, 82]では企業レベルではなく，個人レベルでのコラボレーションに焦点
- 企業のOSS参加に関する先行研究^[40, 90, 93, 98]では動機や参加戦略，ビジネスモデルに焦点

[17] Kevin Crowston, Hala Annabi, James Howison, and Chengetai Masango. 2005. Effective work practices for FLOSS development: A model and propositions. In System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on. IEEE, 197a–197a.

[18] Kevin Crowston, Kangning Wei, James Howison, and Andrea Wiggins. 2012. Free/Libre open-source software development. *Acm Computing Surveys* 44, 2 (2012), 1–35.

[40] Dietmar Harhoff, Joachim Henkel, and Eric Von Hippel. 2003. Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. *Research policy* 32, 10 (2003), 1753–1769.

[78] Param Vir Singh. 2010. The small-world effect: The influence of macro-level properties of developer collaboration networks on open-source project success. *ACM Trans Softw Engineer Methodol* 20, 2 (2010), 6.

[79] Param Vir Singh, Yong Tan, and Nara Youn. 2011. A hidden Markov model of developer learning dynamics in open source software projects. *Information Systems Research* 22, 4 (2011), 790–807.

[82] Xin Tan and Minghui Zhou. 2019. How to Communicate When Submitting Patches: An Empirical Study of the Linux Kernel. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, ArticleArticle 108 (Nov. 2019), 26 pages. <https://doi.org/10.1145/3359210>.

[90] Joel West and Scott Gallagher. 2006. Challenges of open innovation: the paradox of firm investment in open-source software. *R&D Management* 36, 3 (2006), 319–331.

[93] Yuxia Zhang, Minghui Zhou, Audris Mockus, and Zhi Jin. 2019. Companies' Participation in OSS Development-An Empirical Study of OpenStack. *IEEE Transactions on Software Engineering* (2019).

[98] Minghui Zhou, Audris Mockus, Xiujuan Ma, Lu Zhang, and Hong Mei. 2016. Inflow and retention in oss communities with commercial involvement: A case study of three hybrid projects. *ACM Trans Softw Engineer Methodol* 25, 2 (2016).

2. 目的とアプローチ

目的

複数の企業が参加する大規模な
オープンソースエコシステムにおいて
企業間でどのようにコラボレーションするのか
を理解する

アプローチ

OpenStackのコミット履歴や、企業のOpenStackへの
貢献に関する公開情報などを基に、定量・定性的に分析

3. データセット | OpenStackとは



OpenStack

コンピューティング，ストレージ，
ネットワーク等のクラウド
コンピューティング構築に必要な
様々な機能を持つOSSプロジェクト群

OpenStackが企業間コラボレーションの分析に適切な理由

- 数千のリポジトリを持つ大規模エコシステム
- 約10年間のコミット履歴
- 600社以上から貢献を受けている

3. データセット | データ収集

1. コミット履歴の取得

GitHubのミラーリポジトリから2010～2019年のコミット履歴を取得

2. データの前処理

- ボットが投稿したコミットの削除
- 同じ開発者のIDのマージ
- 開発者の所属する組織の特定

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

4. RQ1. 企業はどのように貢献している？ | RQ説明

RQ1.企業はOpenStackにどのように貢献しているのか？

- 動機

企業間コラボレーションが発生する際，各企業がOSS
貢献を行う理由を知るため

- アプローチ

OpenStackに貢献する企業のクラスタリングを行い，
定性的分析によってビジネスモデルのパターンを特定

4. RQ1. 企業はどのように貢献している？ | アプローチ詳細

RQ1. 企業はOpenStackにどのように貢献しているのか？

1. クラスタリング

ノード：企業 or プロジェクト

エッジ：企業がプロジェクトに貢献した場合，エッジを張る
エッジはコミット数で重みづけ

⇒ 企業－プロジェクトの2部グラフを生成
Bi-Louvain法でクラスタリング

2. 定性的な分析

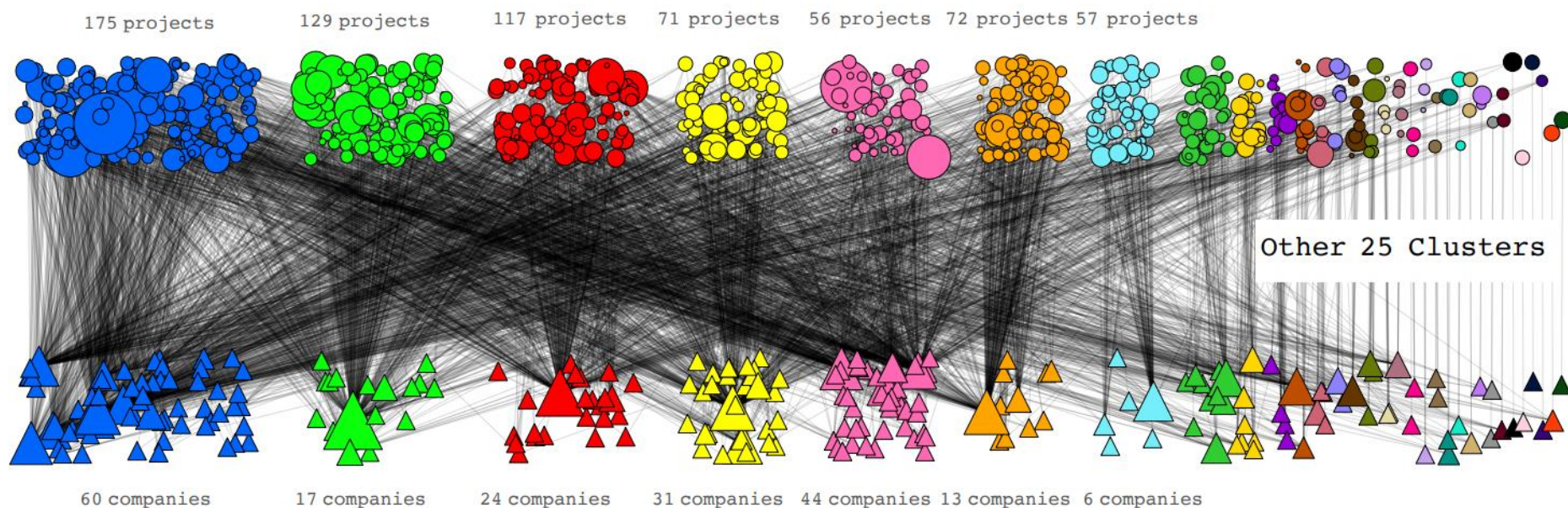
各クラスタ内のコミット数が最も多い企業・プロジェクトを
それぞれ選択

OpenStackや企業の公開情報を基に，著者らによる
テーマ分析によってビジネスモデルをカテゴリ化し
パターンを特定

4. RQ1. 企業はどのように貢献している？ | 結果

クラスタリングの結果

32個のクラスタが出現，モジュラリティ：0.51 (0.3~0.7が目安^[59])



各クラスタ内でコミット数が最も多い企業・プロジェクト
を定性的な分析の対象として選択
ビジネスパターンを特定

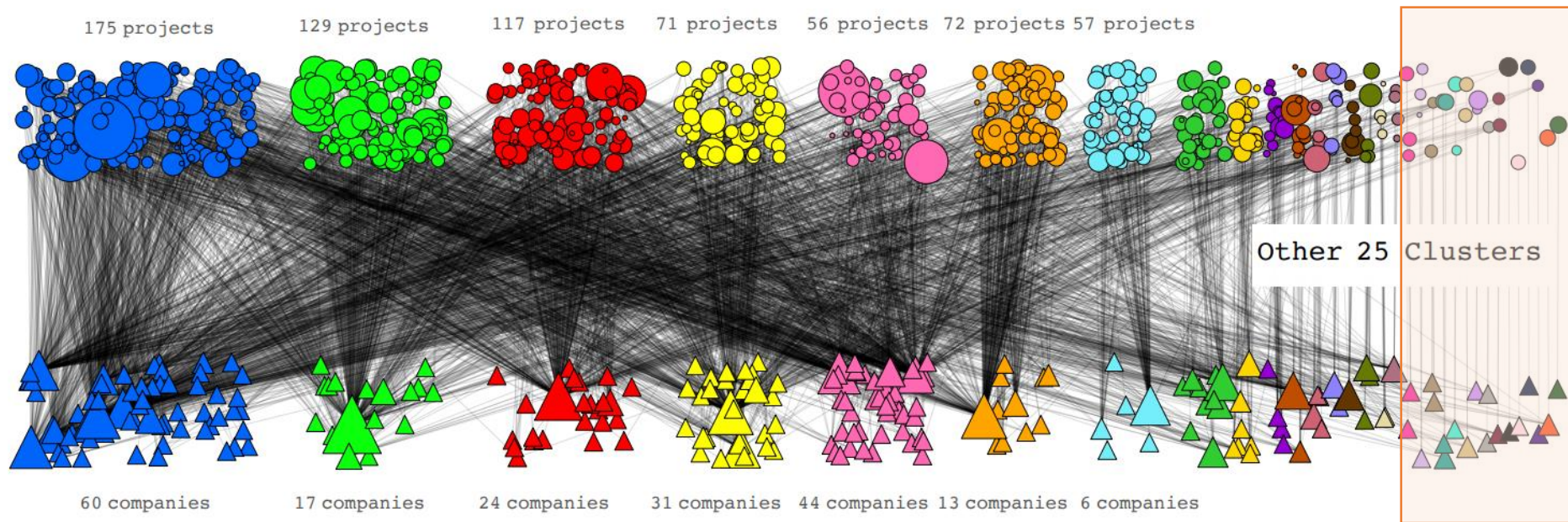
[59] Mark EJ Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. Physical review E 69, 2 (2004), 026113.

4. RQ1. 企業はどのように貢献している？ | 結果

パターン1 (14/32クラスタ)

自社のビジネスをOpenStackと統合するために
プラグインやドライバを開発し貢献

各クラスタの規模は、2~13ノードと小規模

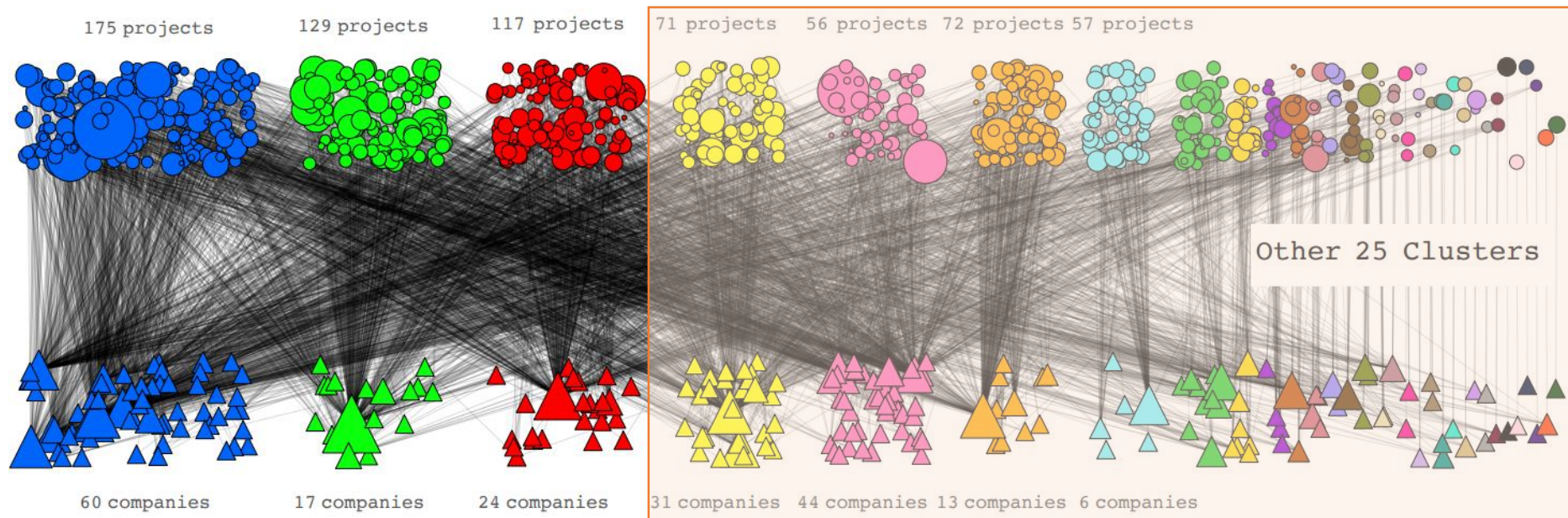


4. RQ1. 企業はどのように貢献している？ | 結果

パターン2 (8/32クラスタ)

OpenStackをベースとしたクラウドサービスを
提供するための、デプロイツールを開発し貢献

各クラスタの規模は、小～中規模

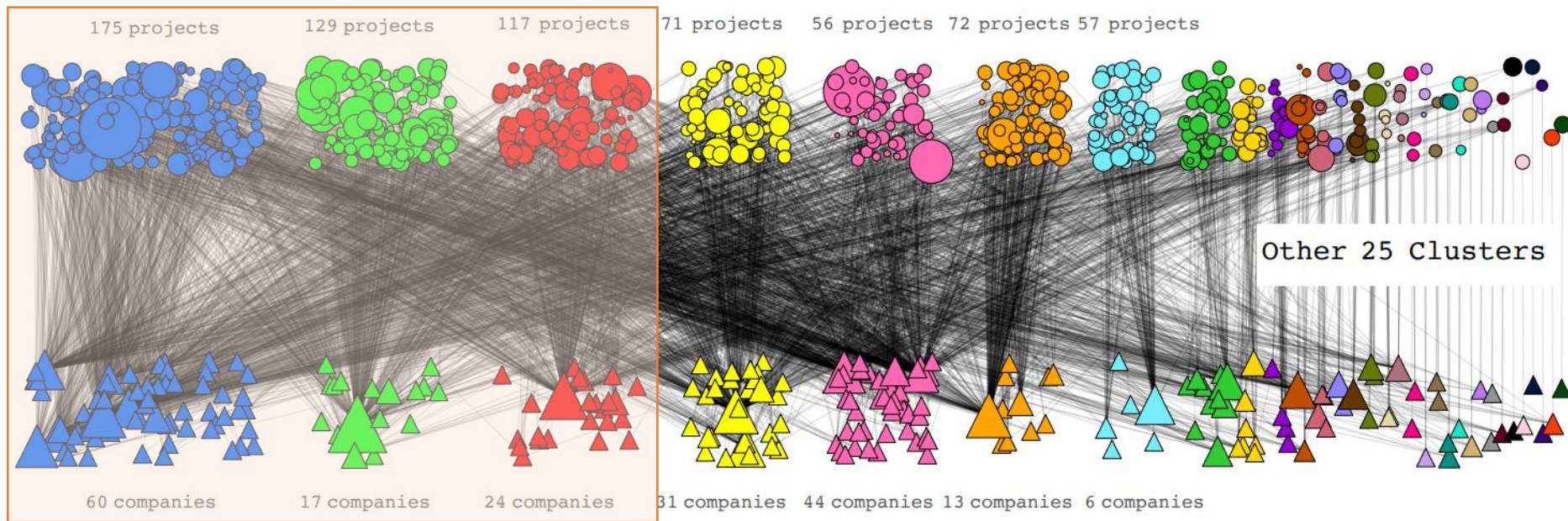


4. RQ1. 企業はどのように貢献している？ | 結果

パターン3 (3/32クラスタ)

それぞれIBM, Huawei, Fujitsuが中心のクラスタ
上記大手企業の委託先を含む、多様なプロジェクトに関与

各クラスタの規模は、100ノード以上の大規模

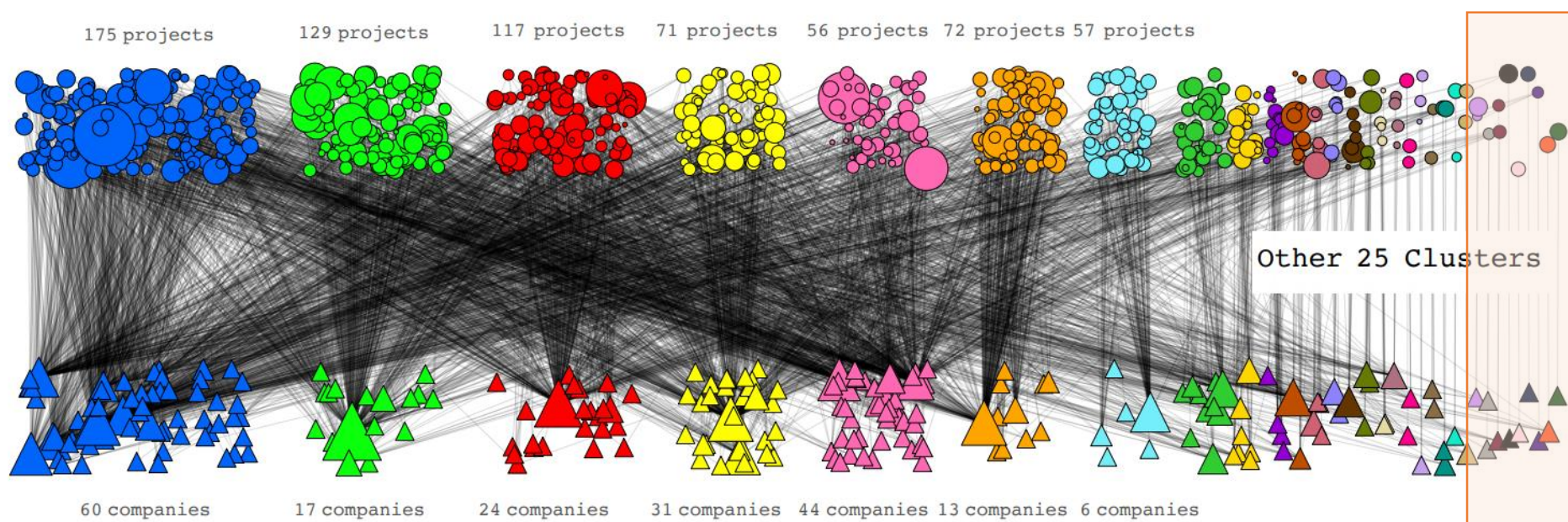


4. RQ1. 企業はどのように貢献している？ | 結果

パターン4 (7/32クラスタ)

新しいサービス等を追加するために，単一の企業が貢献

各クラスタの規模は，10ノード以下の小規模



5. RQ2. 企業間コラボレーションのタイプは？ | RQ説明

RQ2. OpenStack内での企業間コラボレーションのタイプは？

- 動機

企業が同じプロジェクトで共同作業を行う場合，企業間に何らかの関係が存在するのではないかという仮説を確かめる

- アプローチ

同じプロジェクトに貢献した企業のペア計217社のホームページを調査し，ペア企業との共同作業に関連するドキュメントを探す

5. RQ2. 企業間コラボレーションのタイプは？ | アプローチ詳細

RQ2. OpenStack内での企業間コラボレーションのタイプは？

コラボレーションのタイプは以下の3つを想定

ペア企業との共同作業に関するドキュメントが

- ・存在した場合 ⇒ 意図的なコラボレーション
- ・存在しなかった場合 ⇒ 受動的なコラボレーション

- ・そもそもコラボレーションしていなかった場合

⇒ 孤立したコラボレーション

(コラボレーションしなかった/できなかった理由を考察)

5. RQ2. 企業間コラボレーションのタイプは？ | 結果

意図的なコラボレーション

OpenStackを導入したい企業と導入を支援する企業が同じプロジェクトに貢献することが多い

(RQ1のパターン2の貢献を行う企業に見られる)

受動的なコラボレーション

自社ビジネスでOpenStackを使うために貢献する企業間の関係が競合他社である傾向

(RQ1のパターン1の貢献を行う企業に見られる)

孤立したコラボレーション

自社サービス用のプラグイン，ドライバ等，ほかの企業が使わないようなプロジェクトに貢献することが多い

(RQ1のパターン2，4の貢献を行う企業に見られる)

RQ3. OpenStackにおける企業間のコラボレーションは企業の生産性の向上につながるのか

- 動機

先行研究^[12]では，開発者間のコラボレーションは生産性を向上させることが明らかになった
では，企業間のコラボレーションではどうか

- アプローチ

企業間コラボレーションと生産性の関係を回帰分析

[12] Casey Casalnuovo, Bogdan Vasilescu, Premkumar Devanbu, and Vladimir Filkov. 2015. Developer onboarding in GitHub: the role of prior social links and language experience. In Proceedings of the 2015 10th joint meeting on foundations of software engineering. ACM, 817–828.

[22] Daniela Damian and James Chisan. 2006. An empirical study of the complex relationships between requirements engineering processes and other processes that lead to payoffs in productivity, quality, and risk management. IEEE Transactions on Software Engineering 32, 7 (2006), 433–453.

[83] Yong Tan, Vijay Mookerjee, and Param Singh. 2007. Social capital, structural holes and team composition: Collaborative networks of the open source software community. ICIS 2007 Proceedings (2007), 155.

RQ3. OpenStackにおける企業間のコラボレーションは企業の生産性の向上につながるのか

- 各企業の生産性を分析

生産性 = コミット数 / コミットした従業員数
とし、各企業の生産性を求める

Rのboxplot.statsにより外れ値を除去したうえで、
最小，中央，最大値を求める

- 回帰分析

目的変数：生産性

説明変数：RQ1で作成したグラフの次数中心性

コラボレーションが多い企業ほど生産性が高いのか？

企業の生産性の分析結果

N = 602社のうち、

- ・ 最小値：1.0（559社（93%））
- ・ 中央値：6.2
- ・ 最大値：28.5

※生産性 = コミット数 / コミットした従業員数
= 従業員1人当たりのコミット数

企業の生産性の分析結果

- ・ 決定係数：0.19
- ・ 説明変数の係数：11.8（ $P < .05$ ）

⇒ 企業のコラボレーションと生産性の向上には正の関係
ただし決定係数が低いので、ほかに重要な要因があるかも

RQまとめ

- RQ1では，企業がOpenStackの開発に参加する際の4つのパターンを特定
- RQ2では，企業間でコラボレーションが発生する場合の3つのタイプを特定
- RQ3では，企業間コラボレーションによりOpenStackでの企業の生産性が向上することを確認

これらの結果は，企業がOSSエコシステムに参加する際の戦略を考えるときに役立つ