

オープンソース・コミュニティの活性化に寄与する2者関係の探索

芝浦工業大学 大学院
理工学研究科 システム理工学専攻

○新井 健太
中井 豊

1. 概要

本研究ではコミュニティに含まれる開発者同士の2者関係によって構築される環状構造と次数、貢献量を用いてオープンソース・コミュニティの構造を分析し、オープンソース・プロジェクトの活性化との関係性を明らかにする。

2. 研究背景

近年、プロジェクトのソフトウェアをオープンソース化する動きが活発化している。オープンソース化する大きなメリットは、利用者がソースコードを参照できることで、バグの発見や機能の追加に際して、具体的なフィードバックを得られることである。このことから、オープンソース化したプロジェクトではそれを取り巻く利用者の集合、すなわちコミュニティの存在が重要になる[1]とされている。本研究は、オープンソース・コミュニティの構造を分析し、それがプロジェクトの活性化に与える影響を調査する。

3. 関連研究

オープンソースのプラットフォームとして機能している GitHub[3]上のプロジェクトと開発者のネットワーク構造に着目した研究としてはソーシャルコーディングの傾向を分析したもの[2]がある。この研究ではプロジェクト間、開発者間のネットワークの統計指標を算出し、有力なプロジェクトと開発者の発見を行なっている。

4. 研究内容

本研究ではコミュニティに所属する開発者同士の2者関係に注目し、その環状の構造とプロジェクトの活性化の関連性について明らかにする。

プロジェクトに貢献している2人の開発者がいる時、その2人が別のプロジェクトを所有・貢献している場合が考えられる。このような環状の構造を用いれば、コミュニティを

開発者の関係性を通して分析できる。例えば、ある開発者が立てたプロジェクトに対して、別の開発者が常に貢献をしている場合である。これは一方的な協力と呼べる構造だが、自身のプロジェクトに貢献してくれた開発者のプロジェクトに貢献する相互的な協力と呼べる構造も存在する。本研究では、このような2つのプロジェクトと2人の開発者から成る環状構造がプロジェクトに対してどのような影響を与えるのかについて明らかにする。

この場合、理論的には以下の環状構造が考えられる。この内、関係のパターン1は前述の相互的な協力、同じく関係のパターン2は一方的な協力の関係を表している。

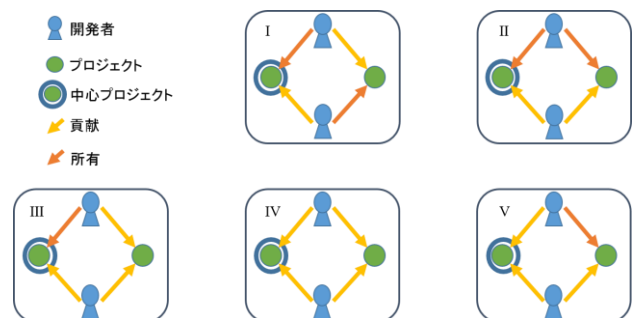


図 1: 環状構造

本研究は、各プロジェクトがそれぞれ幾つの環状構造に囲まれているかに注目することで、プロジェクトとコミュニティの関係を検討する。具体的には環状構造の個数と1つあたりの平均貢献量をそれぞれ算出する。また、プロジェクトの活性化については、Github 上での fork 数を指標として用いる。この fork 数はプロジェクトを複製しての編集している人の数をそれぞれ表している。Github では貢献の際に fork したプロジェクトでの編集を奨励するプロジェクトも多いことから、開発の活性化の指標として妥当だと考えられる。

5. 使用するデータ

使用するデータは Github 社が公開している API[4]を利用して、Github 上でユーザが立

ち上げたプロジェクトや貢献しているプロジェクト、プロジェクトへの貢献量などをプログラムで取得したものである。収集したデータの項目とその総数を以下の表1に示す。

表1：データの概要

種類	名前	属性	データ数
ノード	開発者	Id, 名前, follow 数, follower 数, repo 数, gist 数, 会社, 登録日時	31145
ノード	プロジェクト	id, 名称, fork 数, watch 数, star 数, 開発言語, 作成日時, 更新日時	27935
関係	所有	貢献量	27935
関係	貢献	貢献量	110357

6. データの分析方法

データベースに収集したデータを挿入し、プロジェクトのそれぞれについて環状構造がいくつ含まれるかを算出した。その後、Github 上での fork 数を被説明変数、算出した環状構造の個数と環状構造 1 つあたりの平均貢献量、プロジェクトの次数およびリンク 1 本あたりの平均貢献量を説明変数として重回帰分析を行なった。なお、説明変数・被説明変数ともに対数変換と標準化を行なっている。

7. 分析結果

重回帰分析を行なった結果を表2と表3に示す。なお、決定係数については自由度調整済みの決定係数を記載している。

表2：決定係数

	決定係数	p 値(F 検定)
fork 数	0.319	**

** : 1% 有意

決定係数は有意な値が得られており、およそ 3 割ほどを説明できている。それぞれの相関係数も概ね 1% 有意である。相関係数を個別に見ると、パターン 1 とパターン 2 では環状構造の個数については正の相関だが、平均貢献量については負の相関を示している。逆に、パターン 3 では個数について負、平均貢献量について正の相関を示している。パターン 4 とパターン 5 も真逆でパターン 4 は個数、平均貢献量ともに正の相関を示しているが、

パターン 5 はともに負の相関を示している。

表3：相関係数

	fork 数	
	係数	p 値
切片	-5.01e-13	n.s
Num(P1)	0.0266	*
avC(P1)	-0.0349	**
Num(P2)	0.0708	**
avC(P2)	-0.0276	**
Num(P3)	-0.0667	**
avC(P3)	0.0847	**
Num(P4)	0.0498	**
avC(P4)	0.0545	**
Num(P5)	-0.7098	**
avC(P5)	-0.0690	**
Num(L)	0.8992	**
avC(L)	0.0681	**

※ ** : 1% 有意, * : 10% 有意, n.s. : 非有意

8. 考察と今後の予定

所有を含むパターンは影響が一樣ではなく、扱いが難しいと言える。特に、パターン 5 の負の相関については非常に大きい。これは開発者の貢献量が所有しているプロジェクトに偏ってしまい、中心となっているプロジェクトに注がれないためではないかと考えられる。

今後はデータの総数を増やしての再分析と個数の増加した際の異なる解釈を分析できないか試みる予定である。

9. 参考文献

- [1] 木村誠, 根来龍之: 「営利企業におけるオープンソース形態の取り込み: OSS コミュニティとの関係づけとライセンス形態」, オフィス・オートメーション, 2001, 22. 2: 72-81.
- [2] THUNG, Ferdian, et al: "Network structure of social coding in github". In: Software maintenance and reengineering (csmr)", 2013 17th european conference on. IEEE, 2013, p. 323-326.
- [3] GitHub. (閲覧日: 2016/6/27). <https://github.com>
- [4] GitHub API v3. (閲覧日: 2016/6/27). <https://developer.github.com/v3>