

GitHub におけるソーシャルネットワークの理解： Rockstar を中心として

Understanding the Social Network in GitHub: A Case Study on Rockstars

築山 文香[†]
Ayaka Tsukiyama

山谷 陽亮[†]
Yosuke Yamatani

大平 雅雄[†]
Masao Ohira

1. はじめに

近年のオープンソースソフトウェア（以降では OSS と呼ぶ）は、商用ソフトウェアに引けを取らない機能と品質を備えたものが多くリリースされており、エンドユーザのみならず企業におけるソフトウェア開発においても OSS を盛んに利用するようになった。一方で、広く普及している OSS に致命的な欠陥が存在することもある。OpenSSL における Heartbleed 問題[1]が発覚した際に注目されたように、広く普及している OSS であっても、開発そのものは少数のボランティア開発者から構成される開発コミュニティによって支えられている場合も多い[2]。OSS を多くのユーザが安心して安全に利用するためには、OSS 開発コミュニティの実態を正しく理解しておく必要がある。

このような問題意識を背景として、OSS 開発コミュニティを対象とした分析がこれまで盛んにおこなわれてきた。それらの分析結果を総合すると、(1) 一般の OSS 開発者は平均して 1~1.5 年程度で OSS 開発コミュニティを離脱してしまうこと[3]、(2) 結果的に一部のコア開発者に過度な負担がかかっていること[4]が、多くの OSS 開発コミュニティに共通する問題点として挙げられる。

これらの問題は、現在の OSS 開発コミュニティの運営上の問題に起因すると考えられている。開発中のソースコードはコア開発者のみが自由に書き換えるための特別な権限を持っている。一般開発者が権限を付与される（すなわちコア開発者に昇格する）までには多くの時間を要する。OSS 開発コミュニティにとって信頼に足る有能な人物であるかどうかを見極めるためである。しかしながら、一般開発者が高いモチベーションを維持して OSS 開発コミュニティに貢献し続けることは難しく、短期間での離脱につながっているとされる。

OSS を安定的に保守開発し続けるためには、OSS 開発コミュニティに対する一般開発者の長期的な貢献（とその後のコア開発者への昇格）が必須であるため、長期的貢献を見込めそうな一般開発者を早期に見つけ出すための研究[5]も盛んにおこなわれている。ただしこれまでの研究は、Mozilla や Apache などフルタイム雇用された専任の開発者が存在するような大規模 OSS 開発コミュニティを対象にしたものがほとんどであり、近年主流となりつつあるソーシャルコーディングサイトにおける新しい形態の OSS 開発コミュニティを対象としたものではない。

そこで本研究では、ソーシャルコーディングサイト GitHub[6]における開発者を対象とした分析をおこなう。特に、Rockstar と呼ばれる他の開発者から注目されている開発者を分析し、将来 Rockstar になる可能性の高い開発者を

早期に見出すための予測モデルを構築することを最終目標にしている。そのための第一歩として本稿では、いくつかの事例をあげながら Rockstar の基礎的な分析をおこない、今後の分析方法について議論する。

2. GitHub と Rockstar

GitHub は登録ユーザ数が 1,000 万人を超える現在最も人気のあるソーシャルコーディングサイトの 1 つである。登録開発者は、OSS 開発コミュニティを運営するためのソースコードリポジトリ（すなわち git）や課題管理システムを無料で利用できる。さらに、ソーシャルコーディングと呼ばれる所以[7]となっている、プルリクエストと呼ばれる他の開発者からの貢献（ソースコードの寄贈）を効率的に取り込むための機能や、OSS 開発コミュニティのウォッチ機能・開発者のフォロー機能といった他の OSS 開発コミュニティや開発者の活動を追跡・把握するための仕組みが提供されている。

本研究では、GitHub に登録されている Rockstar と呼ばれる開発者に着目する。Rockstar とは、1,000 以上の開発者にフォローされている注目度の高い開発者を指す。Rockstar は他の多くの開発者が参考にすべき優れた技能や知識を持つと考えることができる。Lee ら[8]の分析によると、Rockstar は GitHub における OSS 開発コミュニティ貢献度が高いだけでなく、Rockstar の活動が手本となるため Rockstar をフォローしている開発者にも良い影響（活動量を高める）を与えるとされている。従って、将来 Rockstar になり得る開発者を早期に見出し OSS 開発コミュニティに定着させることができれば、OSS 開発コミュニティの持続的な発展に大きく寄与すると期待できる。

2. Rockstar の分析

2.1 データセット

本研究では GHTorrent データセット[9]を用いる。GHTorrent データセットは、GitHub API を用いて取得可能な GitHub 上のすべての活動データを保存したものである。GHTorrent データセットを用いることで、Rockstar およびそのフォロワーのすべての活動を時系列に分析できる。本研究の分析対象である Rockstar は、2016 年 6 月 16 日時点において 522 人であった。登録数が 1,000 万人を超える GitHub の開発者の中で極めて一部の開発者であることが分かる。

2.2 Rockstar のフォロワー数

図 1 に 522 人の Rockstar のフォロワーの数を示す。縦軸がフォロワーの数、横軸はフォロワー数の多い順に

[†]和歌山大学, Wakayama University

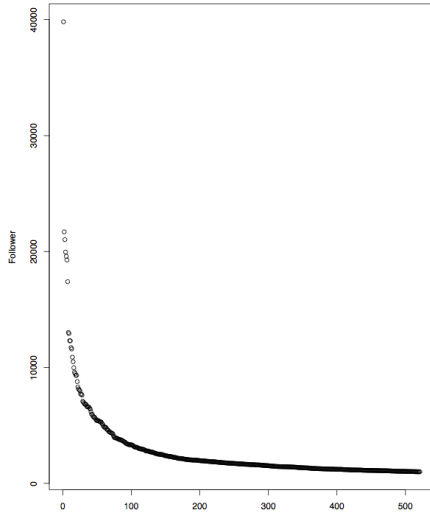


図 1 Rockstar のフォロワー数の分布

Rockstar を並べたものである。

図 1 より、Rockstar のフォロワー数の分布に指数関数的な特徴がみられる。すなわち、最もフォロワー数の多い開発者（Linux 開発者の Linus Torvalds 氏のアカウント）は約 4 万人の開発者からフォローされているのに対して、大多数の Rockstar のフォロワーは数千人の規模である。1 万人以上にフォローされている開発者は 20 名程度であり、偏った分布となっている。一般の多くの SNS でも同様のべき分布が観測されることがあり、ウォッチ機能およびフォロー機能により GitHub においても「ソーシャルな」人間関係が構築されていると推察される。

2.2 Rockstar のフォロワー数の推移

本研究の最終的な目標は将来 Rockstar になり得る有望な開発者を予測するためのモデルを構築することである。したがって、既存 Rockstar がどのようにフォロワーを獲得してきたのか、また、その過程でどのような活動をおこなってきたのかを分析することが必要となる。そこでまず、522 人の Rockstar のフォロワーの獲得の時系列的特徴を 3 つにタイプ分けした。

表 1 Rockstar の分類

タイプ	分類方法
著名人型	登録した月に最も多くのフォロワーを獲得するタイプの開発者。他の著名な OSS 開発コミュニティのリーダーであるなど、GitHub アカウントの登録以前からすでに著名人である可能性の高い開発者。
早世型	登録した月に多くのフォロワーを獲得し、その後に徐々にフォロワーを獲得していくタイプの開発者。注目される OSS 開発コミュニティを開始したと思われる開発者。
晩成型	登録した月とその後の数か月間のフォロワー数に大差がなく、徐々に月ごとのフォロワーが増えるタイプの開発者。GitHub に登録してから注目を集める開発者となり Rockstar になったと考えられる開発者であり、本研究が予測すべき対象。

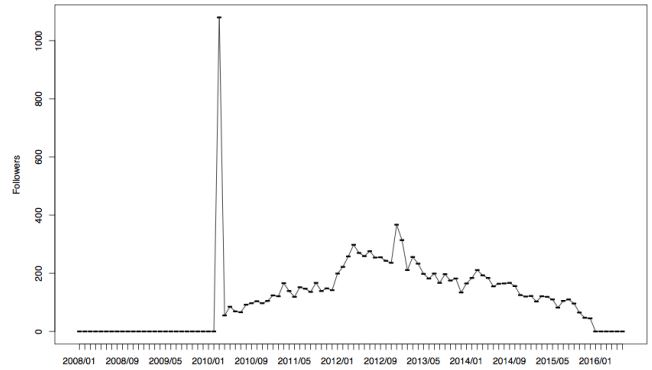


図 2 著名人型 Rockstar のフォロワー獲得推移の例

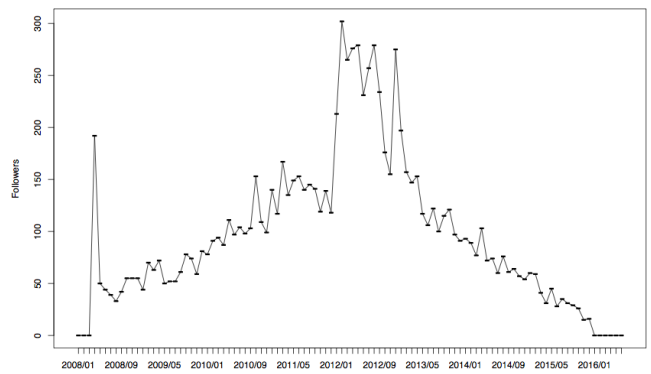


図 3 早世型 Rockstar のフォロワー獲得推移の例

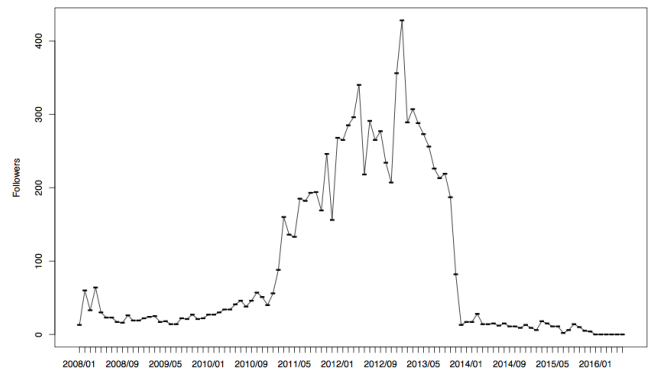


図 4 晩成型 Rockstar のフォロワー獲得推移の例

表 1 に Rockstar の 3 つのタイプと本研究における分類方法を示す。また、図 2~4 にそれぞれのタイプで代表的な Rockstar のフォロワー獲得の推移を示す。

図 2 は著名人型の Rockstar のフォロワー獲得の推移の一例を示すものである。GitHub アカウント登録時点で極めて多くのフォロワーを獲得しており、GitHub アカウント登録時点ですでに著名な開発者である可能性が高い。その後も毎月 2~3 百人規模でフォロワーを獲得しており、Rockstar の中でも数万人規模のフォロワーを獲得している少数の開発者である。

図 3 は早世型の Rockstar のフォロワー獲得の推移の一例を示すものである。GitHub アカウント登録時点でも多くのフォロワーを獲得するが、その後もさらにフォロワー獲得のピークがある。GitHub アカウント登録直後に注目される OSS 開発コミュニティを開始し順調にコミュニティを拡大しているリーダー的存在と推察される。

図4は晩成型のRockstarのフォロワー獲得の推移の一例を示すものである。GitHubアカウント登録時点ではフォロワーは多く獲得しないが、ある時点を境にフォロワー獲得が加速している。GitHubに登録してから注目されるようになったRockstarであると考えることができ、晩成型のRockstarを早期に見い出すことができれば、OSSおよび開発コミュニティの安定的成長を支える開発者として活躍してもらえらる可能性がある。

4. まとめと今後の展開

本稿では、GitHubでRockstarと呼ばれる他のユーザーに注目される開発者を予測するための第一歩として、GitHubのRockstarをGHTorrentデータセットから522人抽出し、Rockstarのフォロワー数の分布に指数関数的偏りが存在することを確認した。また、Rockstarを3つのタイプに分けフォロワー獲得の推移を例示し、本研究が予測の対象にすべきRockstarが晩成型のタイプであることを同定した。

本研究の今後の課題は以下の通りである。

- 早世型・晩成型Rockstarのフォロワー獲得のピーク前後でどのような活動をおこなっていたのかを詳細に分析する。
- 著名人型RockstarがフォローしているRockstarのタイプと活動の内容を詳細に分析する。
- Rockstarが参加しているOSS開発コミュニティは、活発に活動し注目を集めている（ウォッチされている）「Rockstarコミュニティ」であるかどうかを分析する。
- Rockstarコミュニティには、Rockstarは初めから参加しているのか、Rockstarコミュニティの成長とともにRockstarとなるのか、Rockstarコミュニティとなった後から参加するのかについて分析する。
- Rockstarがフォローした開発者はロックスターになりやすいのかどうかを詳細に分析する。
- 上述の分析を通じて、特に晩成型RockstarのRockstarになる前の活動の特徴を抽出し、Rockstar候補者予測モデルを構築する。

謝辞

本研究の一部は、文部科学省科学研究補助金（基盤（C）：15K00101）による助成を受けた。

参考文献

- [1] The Heartbleed Bug: <http://heartbleed.com>
- [2] 伊原彰紀, 大平雅雄, オープンソースソフトウェア工学, コンピュータソフトウェア, Vol.33, No.1, pp.28-40, 2016.
- [3] C. Bird, A. Gourley, P. Devanbu, A. Swaminathan and G. Hsu, Open Borders? Immigration in Open Source Projects, Proceedings of the Fourth International Workshop on Mining Software Repositories (MSR' 07), pp.1-8 (CD-ROM, No.6), 2007.
- [4] A. Mockus, R.T. Fielding and J.D. Herbsleb, Two case studies of open source software development: Apache and Mozilla, *ACM Transactions on Software Engineering and Methodology (TOSEM)*, No.11, No.3, pp.309-346, 2002.

- [5] M. Zhou and A. Mockus, What make long term contributors: willingness and opportunity in OSS community, Proceedings of the 34th International Conference on Software Engineering (ICSE '12), pp.518-528, 2012.
- [6] GitHub: <https://github.com>
- [7] L. Dabbish, C. Stuar, J. Tsay, and J.D. Herbsleb, Social coding in GitHub: transparency and collaboration in an open software repository, Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12), pp.1277-1286, 2012.
- [8] M.J. Lee, B. Ferwerda, J. Choi, J. Hahn, J.Y. Moon and J. Kim, GitHub developers use rockstars to overcome overflow of news, Extended Abstracts on Human Factors in Computing Systems (CHI '13), pp.133-138, 2013.
- [9] GHTorrent : <http://ghtorrent.org>