

Open Borders?

Immigration in Open Source Projects

Christian Bird, Alex Gourley, Prem Devanbu,
Anand Swaminathan, Greta Hsu

FOURTH INTERNATIONAL WORKSHOP ON MINING SOFTWARE
REPOSITORIES (MSR'07)

B3 片桐直哉

背景

オープンソースソフトウェアはボランティアの人が集って開発される

- 開発者のコアチームが、リポジトリへのコミット権限を持つ
- コアチームには何らかのプロセスを経て選出される

OSSプロジェクトの継続力は惹きつける能力に依存している

- 開発者を引き入れたり、保持し続ける必要がある
- コアチームへの参加を募る必要がある

参加したがる人はこういったプロセスを辿ればよいのか？

関連研究

プログラミングは豊富な知識を必要とする

- 経験を積んだプログラマでさえ更なる技術の会得が強いられる
- 新規参加者への壁

メーリングリストなどでやり取りを継続して行う

- コア開発者になった人は一般的に行っていた

メーリングリストやコミットログ

- OSSプロジェクトにおける主要なメディア

概要

プロジェクトへの技術的貢献

- プロジェクトの成功に向けどれだけ身を委ねたか
- どれくらいパッチ投稿するか

スキルレベル

- どれくらいの技術、知識が必要とされるか

個人の評価

- コミュニティにおける個人の評価

傾向

メーリングリストが活発な人は確実に減っていく

- 無給のボランティアが長期間モチベーションを維持するのは困難

知識とスキル

- 一回目のパッチ投稿はメーリングリストに参加してから2か月目に見られやすい
- 技術習得に掛かる時間

個人の評価

- Postgresプロジェクトにおいては3-4年で投稿数が減る傾向

個人の評価の指標-*indegree*

indegree

- コミュニティにおけるネットワークの分析
- ある開発者の評判の指標
- Aさんが投稿したメッセージを、興味深いとした人の割合

*indegree*と、パッチ投稿数には相関があることがわかっている

- Postgresプロジェクトにおいては、3-4年で*indegree*は低下する

仮説

仮説1：プロジェクトのコアチームに参加する見込みはある期間まで増加し、以降は下降していく

仮説2：パッチの投稿のようなスキルの活用はコアチーム参加にあたり良い影響をもたらす

仮説3：コミュニティ内でのステータスはコアチーム参加にあたり良い影響をもたらす

分析

メトリクス

- メールの総数
- メーリングリストに参加している非コアチーム参加者の数
- 最終的にコアチームに参加できた人の数
- 過去にパッチが受理されたことがあるかどうか
- *indegree*
- パッチの受理された割合
- コミュニティ内の開発者数
- メッセージを送った数

データセット(1)

Postgresプロジェクト(10年間)

- メール総数 : 169,118
- 非コアチーム開発者数 : 3,283
- コアチーム入りを果たした開発者数 : 19

Postgres	平均	標準偏差	最小値	最大値
パッチ投稿	0.04	0.20	0	1
<i>indegree</i> の 最頻値	0.0024	0.0065	0	0.18
パッチ採用率	1.27	9.78	0	100
コミュニティ内 開発者の総数	13.92	4.14	1	20
一月当たりの メールの総数	13.12	63.55	1	2370

データセット(2)

Apache Web Serverプロジェクト(5年間)

- メール総数 : 50,170
- 非コアチーム開発者数 : 1,445
- コアチーム入りを果たした開発者数 : 30

Postgres	平均	標準偏差	最小値	最大値
パッチ投稿	0.10	0.30	0	1
<i>indegree</i> の 最頻値	0.026	0.049	0	0.087
パッチ採用率	1.13	8.98	0	100
コミュニティ内 開発者の総数	46.24	9.76	3	57
一月当たりの メールの総数	7.34	18.18	1	273

データセット(3)

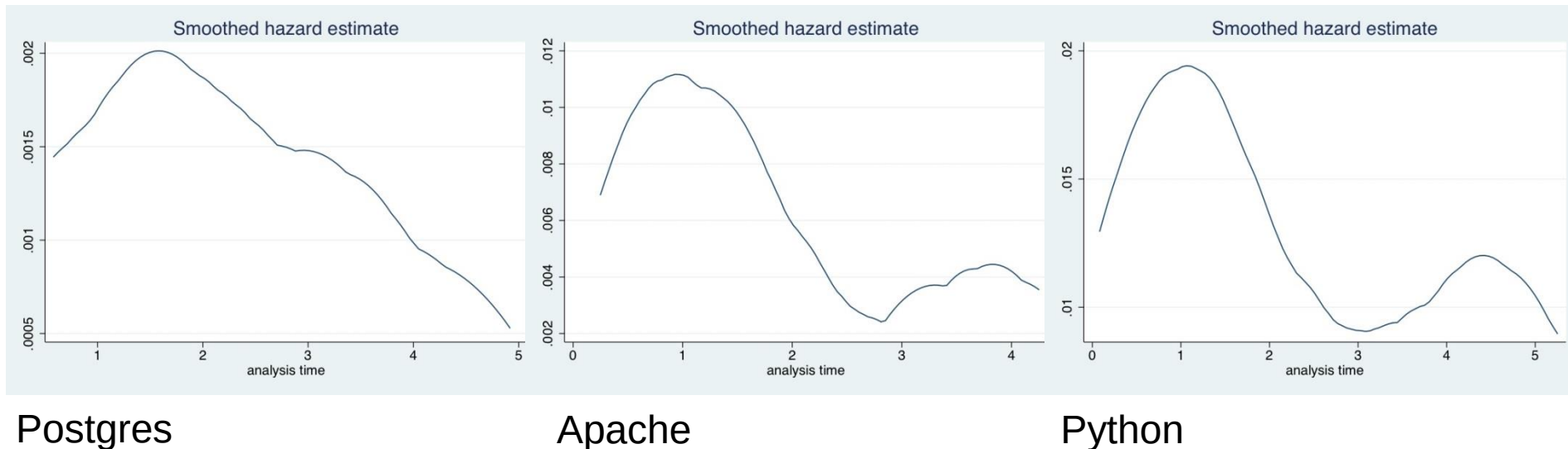
Pythonプロジェクト(7年間)

- メール総数 : 45,216
- 非コアチーム開発者数 : 1,320
- コアチーム入りを果たした開発者数 : 62

Postgres	平均	標準偏差	最小値	最大値
パッチ投稿	0.04	0.2	0	1
<i>indegree</i> の 最頻値	0.0098	0.032	0	0.70
パッチ採用率	0.62	6.98	0	100
コミュニティ内 開発者の総数	58.69	6.98	0	79
一月当たりの メールの総数	13.87	54.44	1	1648

分析結果(1)

生存曲線



縦軸：コアチームに参加できた人数の割合

横軸：経過時間(年)

仮説1

プロジェクトのコアチームに参加する見込みはある期間まで増加し、以降は下降していく

Apache, Postgresプロジェクトに関しては支持された

Pythonプロジェクトは支持されなかった

- 他の二つより歴史が浅いため伸び白があると考えられる

分析結果(2)

影響度	Apache	Postgres	Python
<i>indegree</i>	61.21	72.2	5.49
パッチ投稿の有無	2.71	0.98	2.09
パッチ採用率	0.15	0.0039	0.0062
メール総数(月)	0.014	0.0158	0.0041
開発者数	0.14	1.09	0.14

プロジェクトに変数を与えた影響度は、どのプロジェクトにおいても *indegree* による影響が最も大きかった

仮説2

パッチの投稿のようなスキルの活用はコアチーム参加にあたり、良い影響をもたらす

少なからず良い影響はある

影響度	Apache	Postgres	Python
パッチ投稿の有無	2.71	0.98	2.09

Apache, Pythonプロジェクトにおいて、支持された

Postgresプロジェクトにおいては、あまり支持されなかった

- Postgresプロジェクトは、コアメンバーであろうとなかろうと、パッチ投稿が盛んであると推測される

仮説3

コミュニティ内でのステータスはコアチーム参加にあたり良い影響をもたらす

影響度	Apache	Postgres	Python
<i>indegree</i>	61.21	72.2	5.49

どのプロジェクトにおいても最もコアチーム参加にあたって重要な変数であった

仮説3は全プロジェクトにおいて支持された

まとめ

仮説1：プロジェクトのコアチームに参加する見込みはある期間まで増加し、以降は下降していく

仮説2：パッチの投稿のようなスキルの活用はコアチーム参加にあたり良い影響をもたらす

仮説3：コミュニティ内でのステータスはコアチーム参加にあたり良い影響をもたらす

上記3つの仮説は程度はあれども、支持されることが確認できた

コアチームに参加するにあたって最も重要なことはプロジェクト参加者の関心を得ることであろう