

# **Which Contributions Predict**

# **Whether Developers Are Accepted Into GitHub Teams**

Justin Middleton, Emerson Murphy-Hill, Demetrius Green,  
Adam Meade, Roger Mayer, David White, Steve McDonald

# 背景

- ▶ OSSが発展するにつれて、プロジェクトの維持管理が重要になる
- ▶ チームにどのようにメンバーが参入するか知ること  
は、成功するチームのプロセスを知るために必要となる

- ▶ RQ1 チームに参加する(しない)メンバーでインタラクションがどう異なっているか
- ▶ RQ2 あるチームからの独立メンバーの活動や資質はチームに参加するかどうかに影響するか
- ▶ RQ3 チームに入る人の数はどのような活動や資質に影響されるか

# データセット

- ▶ GHTorrentから開発者データを収集
  - ◆ 2015年8月~2017年1月までを対象
  - ◆ 2015年のスナップショット以前にプルリクしているユーザのプロジェクトに限定

# Membershipについて

- ▶ GitHubにはCollaboratorやMemberなどユーザごとにmembershipが与えられる
- ▶ GHTorrentだけではmembershipの対応づけができないため手動で対応関係付け
  - ◆ 2015年ではOwner,2017年ではContributor

# Membershipの決め方

- ▶ membershipステータスが非公開の開発者がいる
  - ◆ メンバのみが出来る行動(プルリクのマージ)をした開発者をメンバと定義する

**Table 1: Movement in Teams**

|      |              | 2015       |              |       |
|------|--------------|------------|--------------|-------|
|      |              | [No label] | Collaborator | Owner |
| 2017 | [No Label]   | -          | Leave        | Leave |
|      | Contributor  | -          | Leave        | Leave |
|      | Collaborator | Join       | -            | -     |
|      | Member       | Join       | -            | -     |

# データの前処理

- ▶ メンバでないチームとのインタラクションを集計するため、すでにメンバ関係となっている場合は除いている
- ▶ 2015年以降にインタラクションのあるユーザのみ
- ▶ -botで終わるアカウントを取り除く

# 分析

- ▶ ロジスティック回帰を利用
  - ◆ 説明変数是对数変換
- ▶ 全データを使うのではなく2つに分けて予測
  - ◆ 同チームから開発者2人(参加, 不参加)
  - ◆ 同開発者の2チーム(参加, 不参加)

# 前処理後のデータ

**Table 2: Data Parameters**

| <b>Data Subset</b>                         | <b>N</b> |
|--------------------------------------------|----------|
| Outsider Relationships in 2015             | 252,243  |
| Relationships After Cleaning               | 247,589  |
| Relationships with Contributions In 2015-7 | 64,270   |
| Nonjoining Relationships                   | 61,096   |
| Joining Relationships                      | 3,174    |
| Unique Developers                          | 47,352   |
| Unique Teams                               | 11,095   |
| Analysis 1—Paired Developers by Teams      | 2,440    |
| Analysis 2—Paired Teams by Developer       | 1,036    |

# 結果(効果あり)

**Table 3: ODDS RATIO (OR) PER UNIT INCREASE OF EACH METRIC**

|                                                                                                                                              | $\log_2(\text{Metric})$           | Analysis 1 |              | Analysis 2 |              |
|----------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|------------|--------------|------------|--------------|
|                                                                                                                                              | Intercept                         | 0.08***    | [0.04, 0.14] | 0.04***    | [0.01, 0.14] |
|                                                                                                                                              | Statistically Significant Results |            |              |            |              |
| RQ1: What types and qualities of project interactions distinguish between developers who join an OSS team and developers who remain outside? | Pull Requests (PRs)               | 1.58***    | [1.42, 1.77] | 1.86***    | [1.50, 2.33] |
|                                                                                                                                              | Issue Comments                    | 1.04       | [0.95, 1.14] | 1.20*      | [1.04, 1.39] |
|                                                                                                                                              | Comments on Other PRs             | 1.20***    | [1.09, 1.32] | 1.27**     | [1.07, 1.51] |
|                                                                                                                                              | Tenure (Days)                     | 1.08*      | [1.00, 1.16] | 1.14*      | [1.01, 1.29] |
| RQ2: What activities or qualities of a developer independent from a specific team influence whether the developer joins that team?           | Commits to Personal Projects      | 1.07*      | [1.00, 1.14] |            |              |
|                                                                                                                                              | Followers                         | 1.15**     | [1.05, 1.26] |            |              |
| RQ3: What activities or qualities of a team influence whether the team accepts more people in?                                               | Projects in Team                  |            |              | 1.20**     | [1.06, 1.35] |
|                                                                                                                                              | Commits in Team                   |            |              | 0.84**     | [0.74, 0.95] |
|                                                                                                                                              | PRs in Team                       |            |              | 1.30**     | [1.07, 1.57] |
|                                                                                                                                              | Contributors to Team              |            |              | 1.21*      | [1.00, 1.45] |
|                                                                                                                                              | Watchers, Forkers in Team         |            |              | 0.84**     | [0.73, 0.95] |

Analysis 1—Paired Developers by Teams

Analysis 2—Paired Teams by Developer

# 結果(効果なし)

| Statistically Insignificant Results |                                    |      |              |      |              |
|-------------------------------------|------------------------------------|------|--------------|------|--------------|
| RQ1                                 | Issues                             | 1.08 | [0.96, 1.22] | 0.93 | [0.77, 1.13] |
|                                     | Commit Comments                    | 1.11 | [0.93, 1.34] | 0.94 | [0.71, 1.25] |
|                                     | Comments on Own PRs                | 1.00 | [0.92, 1.07] | 1.05 | [0.91, 1.22] |
|                                     | Team Projects Forked               | 1.11 | [0.96, 1.28] | 1.05 | [0.79, 1.38] |
|                                     | Team Projects Watched              | 0.93 | [0.82, 1.06] | 1.07 | [0.88, 1.31] |
| RQ2                                 | Personal Projects                  | 0.99 | [0.90, 1.09] |      |              |
|                                     | Other PRs                          | 1.01 | [0.92, 1.09] |      |              |
|                                     | Other Issues                       | 0.95 | [0.88, 1.04] |      |              |
|                                     | Other Commit Comments              | 0.98 | [0.92, 1.06] |      |              |
|                                     | Other Issue Comments               | 0.98 | [0.90, 1.07] |      |              |
|                                     | Other Comments to Own PRs          | 1.06 | [0.98, 1.15] |      |              |
|                                     | Other Comments to Other PRs        | 1.00 | [0.94, 1.07] |      |              |
|                                     | Following                          | 0.99 | [0.94, 1.05] |      |              |
| RQ3                                 | Watchers, Forkers for Own Projects | 0.96 | [0.90, 1.01] |      |              |
|                                     | Issues in Team                     |      |              | 1.11 | [0.88, 1.40] |
|                                     | Commit Comments in Team            |      |              | 1.09 | [0.99, 1.20] |
|                                     | Issue Comments in Team             |      |              | 0.86 | [0.71, 1.06] |
|                                     | PR Comments to Org                 |      |              | 0.97 | [0.84, 1.13] |

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

Analysis 1—Paired Developers by Teams

Analysis 2—Paired Teams by Developer

# 考察

- ▶ PRが重要→PR周りのインタフェースを整備すると良さそう
- ▶ Independent userについて
  - ◆ follower数が重要となり， 別研究とも一部関わる
    - follower数と開発者の貢献が承認されるか
    - follower数とcoding rockstar
  - ◆ follower数の重要性の裏付けになる

# Implications

- ▶ Community and Website Design
  - ◆ 重要とわかったメトリクスを元にデザインすると新規参入しやすくなるかも
- ▶ Educators
  - ◆ 学生にどこに注力すべきか教えられる
  - ◆ 特定言語やツールの使い方を教えるのとは違うため  
今回の結果はパターンの明示として役立つ

# まとめと所感

## ▶ まとめ

- ◆ GitHubでのチーム参入について分析した
- ◆ PRやfollower数の影響が高かった

## ▶ 所感

- ◆ デザイン/教育方面での考察が面白い