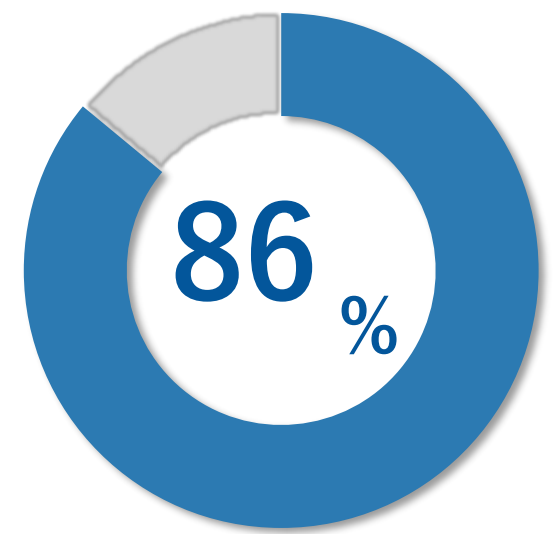


# 神話：OSSコミュニティの参加者はトップ開発者だ

草間 紀輝 (九州大学),  
近藤 将成 (九州大学), 斎藤 忍, 飯村 結香子(NTT), 鵜林 尚靖, 亀井 靖高 (九州大学).

## 背景：OSS開発者の需要の高まり [1]



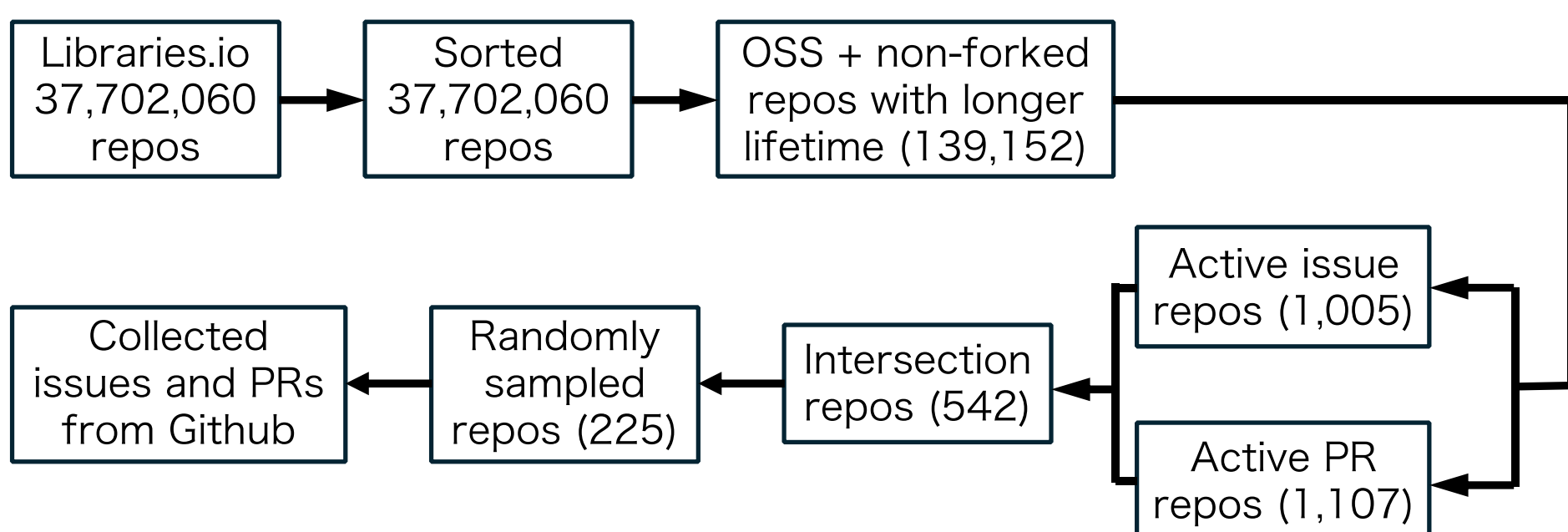
86%の採用マネージャーが  
オープンソース人材の採用が  
優先事項と述べている

### 神話 “OSSコミュニティの 参加者はトップ開発者だ”

[1] [https://training.linuxfoundation.org/wp-content/uploads/2022/06/OpenSourceJobsReport2022\\_FINAL.pdf](https://training.linuxfoundation.org/wp-content/uploads/2022/06/OpenSourceJobsReport2022_FINAL.pdf)

## 分析対象 [2]

リポジトリ：225件      開発者：1,008,257人



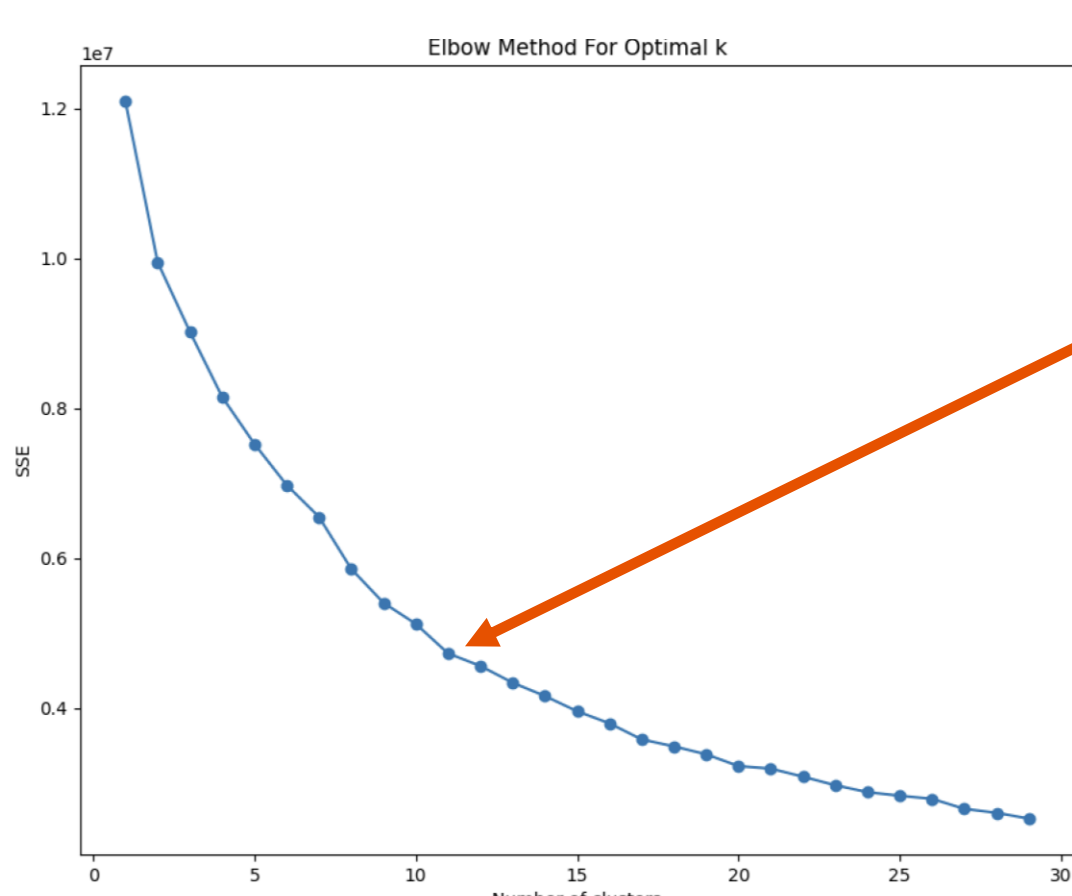
[2] M. Kondo et al., "Towards Better Online Communication for Future Software Development in Industry," 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), Torino, Italy, 2023, pp. 1619-1624, doi: 10.1109/COMPSAC57700.2023.00250. keywords: (Industries; COVID-19; Companies; Programming; Remote working; Data mining; Task analysis; mining software repository; online communication; open-source software; industrial development).

## 調査方法：12の特徴量を定義して クラスタリング

### 特徴量

- coding関連
  - commit数
  - 平均commit間隔
  - PR回数
  - 平均PR間隔
- issue関連
  - issue貢献度
  - issue commentの平均長
  - 平均issue作成間隔
- code review関連
  - code review貢献度
  - code reviewの平均長
  - 平均code review間隔
- 総合指標
  - 総合貢献度
  - コーディング/ディスカッションバランス

### クラスタリング



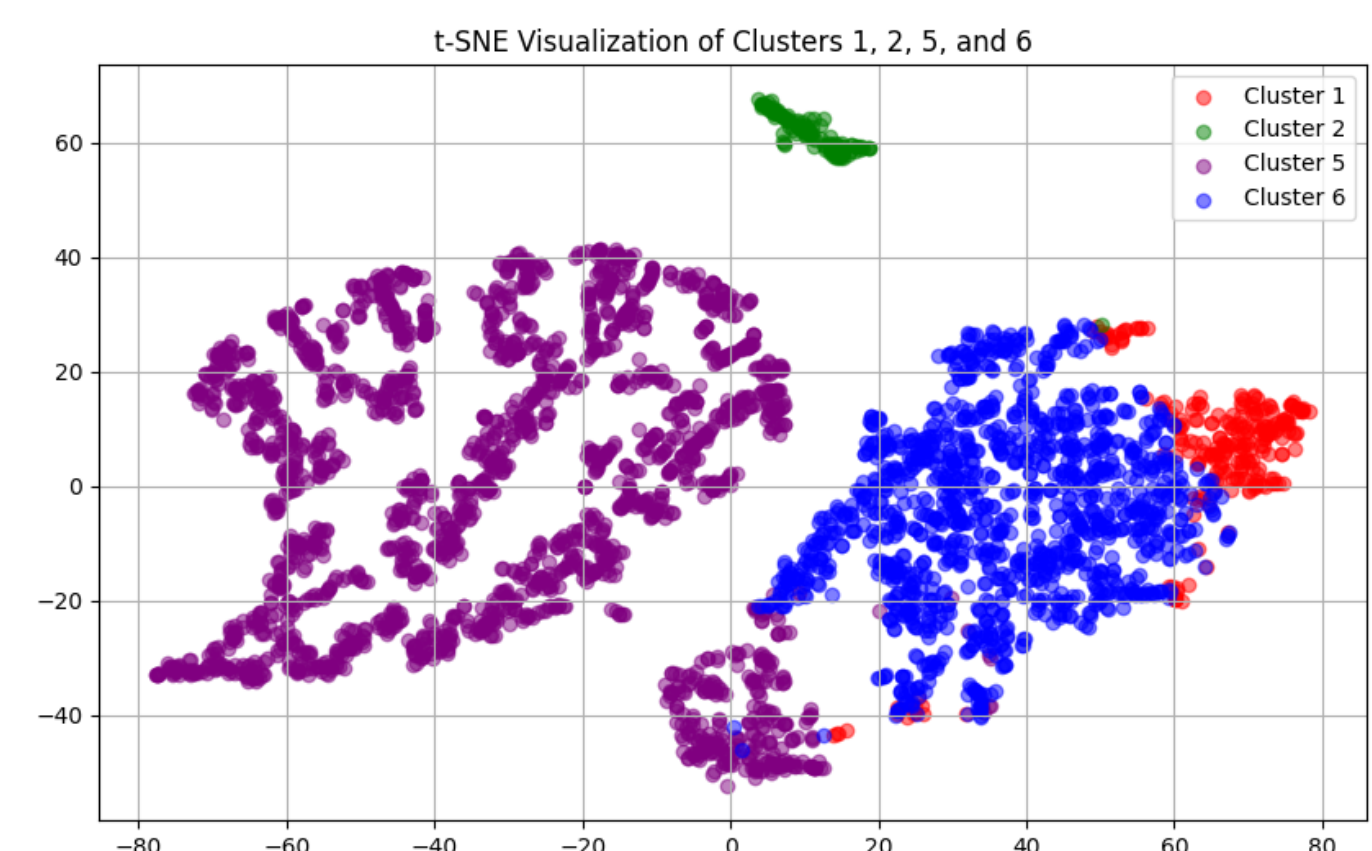
エルボー法によりクラ  
スタ数を11に決定

k-means法によりクラ  
スタリングを行う

## 分析

クラスタリング結果をもとに各クラスの開発者を命名

1. ボランティア：全活動量が平均未満. 全体の95%
2. 一般ユーザ：issue作成間隔が大きい
3. レビュー貢献者：code reviewの平均長が大きい
4. イシュー貢献者：issueコメント平均長が大きい
5. ゆったり開発者：活動間隔が大きい
6. バランスド開発者：バランスよく平均以上の活動
7. エピソード開発者：平均commit間隔が大きい
8. コミッター：commit活動に偏っている
9. 準コミッター：活動傾向はコミッターと似ており  
活動量が小さい
10. コア開発者：総合貢献度が最も大きい  
その他の特徴量も高い値
11. 準コア開発者：活動傾向はコア開発者と似ており,  
活動量が小さい



## 調査結果：トップ開発者ばかりではない！

**発見1:** 多数の名もなきボランティアによる貢献  
に支えられている

ボランティア開発者**95%** vs コア開発者**0.02%**

**発見2:** 役割はバラエティに富んでいる

- コーディングを主な活動とする開発者
- issueやreviewを主な活動とする開発者
- 活動ペースがゆっくりな開発者

➡神話は誤り！

## 貢献：期待する人材の発見に寄与

シナリオ：企業の採用

従来：OSSコミュニティへの参加経験の有無

今後：本研究で示した特徴量を用いることで,  
期待する開発者の発見に寄与