

R community on GitHub

Aims: explore the structure of the R open-source community on GitHub; study how different subcommunities treat newcomers

Data:

- **4419** R packages stored on GitHub,
- **>700k** direct code changes (commits),
- **5.9k** code revisions (pull requests) for top-50 packages

Methods:
descriptive statistics,
network analysis

About R:

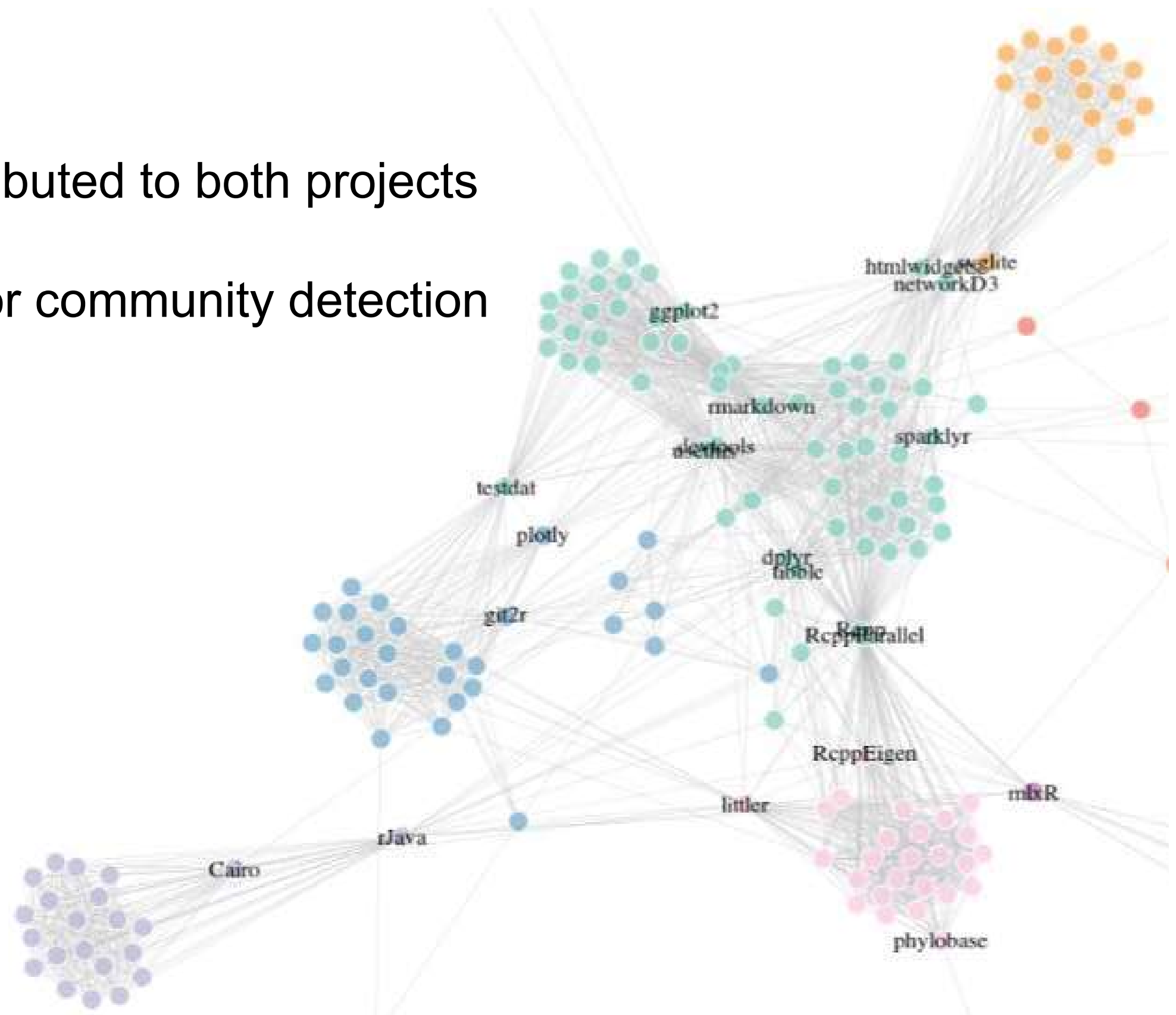
- ranked 12 among the popular languages on GitHub (<http://githut.info/>)
- moved up to 5th place in IEEE language rankings for scientific research in 2017

Projects network

Nodes — projects

Links — someone contributed to both projects

Fast-Greedy algorithm for community detection



Network of top developed packages

Pull requests (code revisions)

Core Team – owner of the package repository on R or those, who have the right to accept the pull request

Middle developers – made 10 and more pull requests

Newcomers – less than 10 pull requests

Pull requests:

59% from the core team

External developers:

19% from the middle

22% from newcomers

73% pull requests were merged out of all closed

Pull requests (code revisions)

Different 'treatment' of pull requests:

80% of pull requests from the core team were merged

72% from the middle developers

58% from the newcomers

This varies for different clusters of packages:

