IETF Hackathon: BigBang

IETF 111
July 19-23, 2021
Online

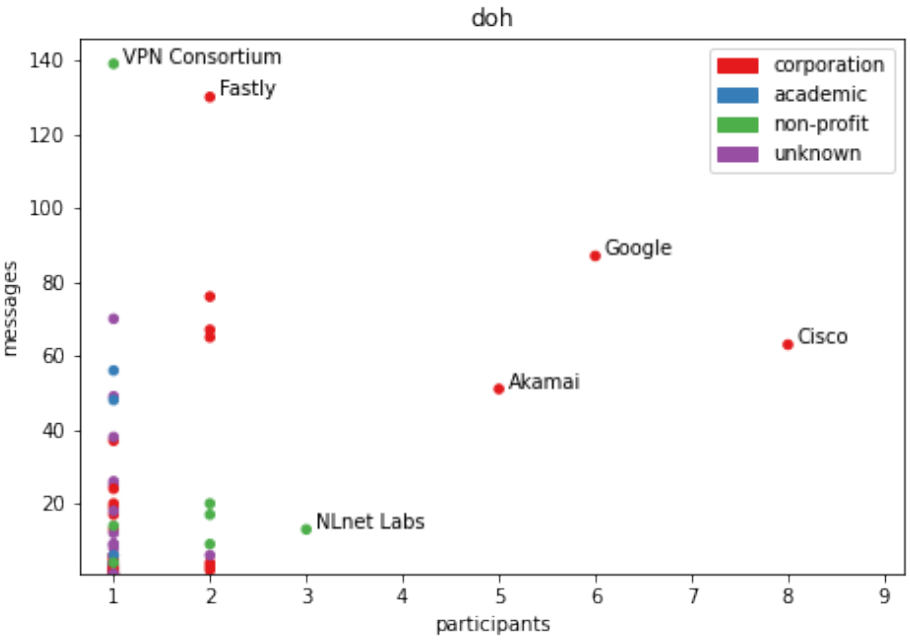
What is BigBang?

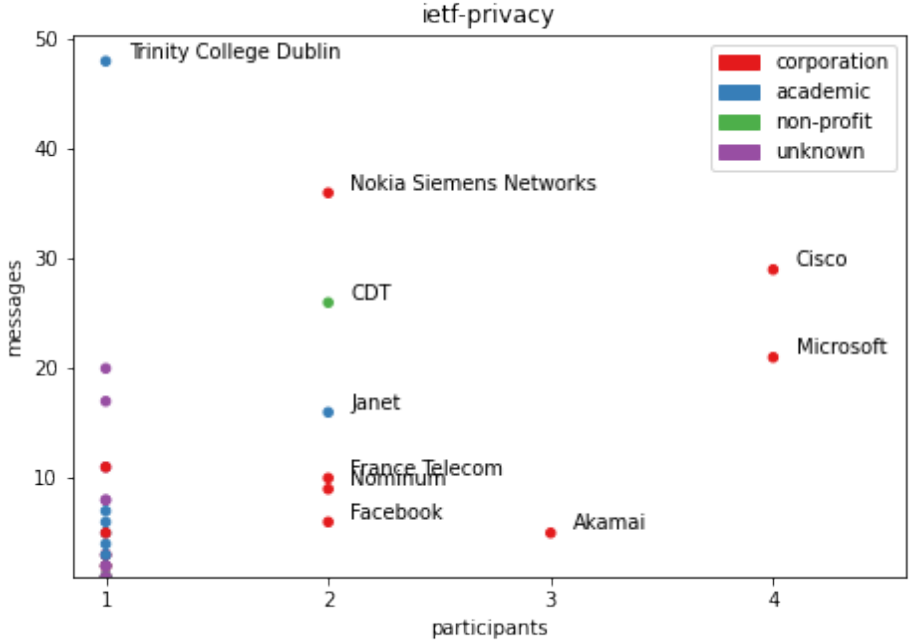
- A scientific toolkit for studying collaborative communities
- Data sources: Email, Git, IETF DataTracker, ListServ, Mailman, Pipermail ...
- Data science tools: using SciPy stack
  - entity resolution
  - social network analysis
  - natural language processing
  - etc.

Hackathon Plan

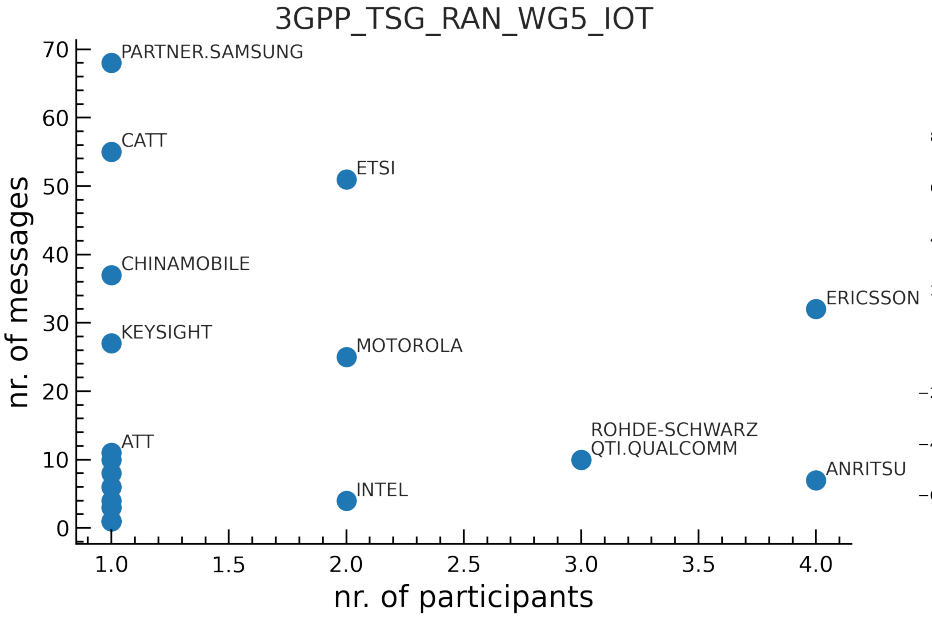
- Add new data sources
  - ingesting mailing lists from other fora: 3GPP, IGF
  - connecting Datatracker and GitHub to mailing list data
- Explore organizational affiliation and email domains

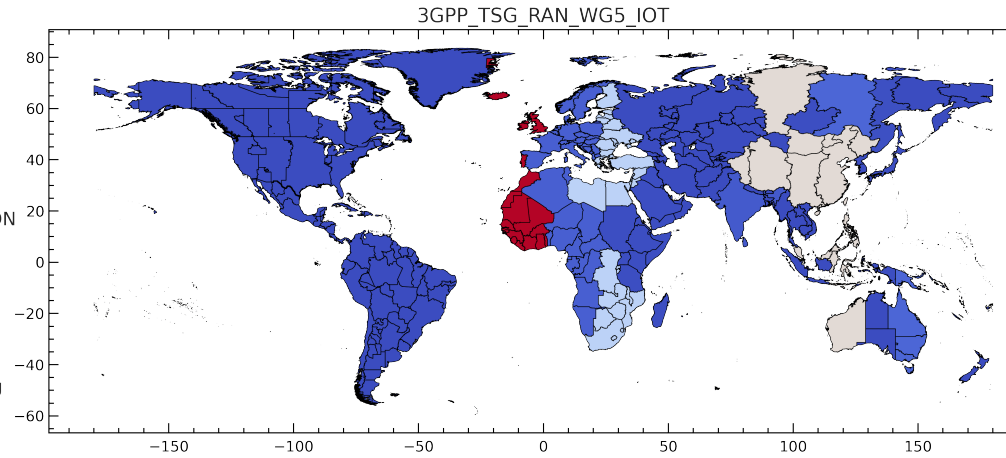
Participation by organizational affiliation and sector



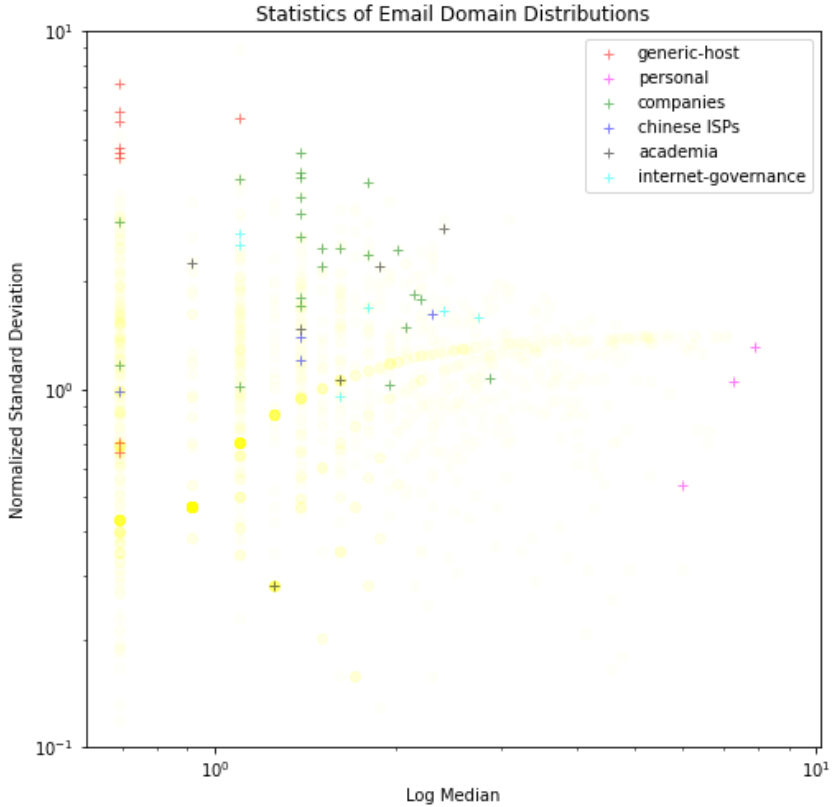


Participation by organizational affiliation and sector





Distribution of participants by email domain



What we learned

- Lessons learned (or re-learned)
  - all data is messy
  - organizational data from Datatracker is promising (Github less so), including draft authorship and meeting attendance
  - useful to track affiliation changes over time
  - remote debugging more challenging than over-the-shoulder help

Wrap Up

Team members:

- Nick Doty
- Sebastian Benthall
- Christoph Becker
- Riccardo Nanni

https://datactive.github.io/bigbang/

Send us your research questions!