BigBang Update
//
Using
Complex Systems Analysis
to Identify
Organizational Interventions

Sebastian Benthall, PhD
Information Law Institute
NYU School of Law

What is BigBang

- A scientific toolkit for studying collaborative communities
- Data sources: Email, Git repositories, IETF DataTracker, ListServ, …
- Data science tools: using Scientific Python stack
  - Entity resolution for names and organizations
  - Social network analysis
  - Natural language processing on message content
  - Time series analysis
  - Information extraction…

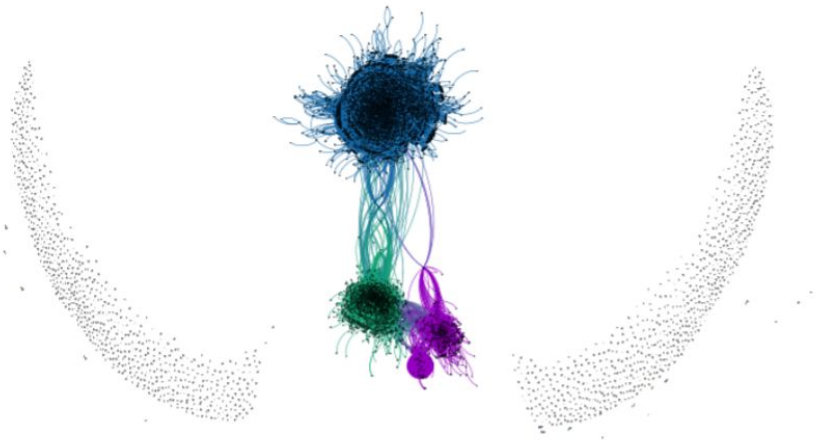
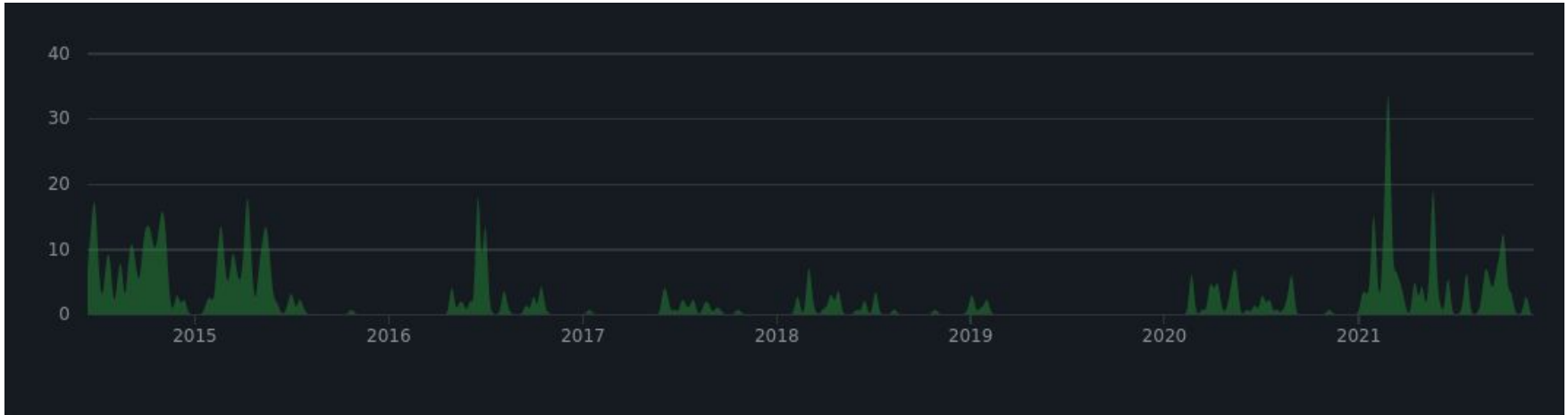


Fig. 1: *Interaction graph of all participants across all mailing lists explored in this study, rendered with [Gephi]. The large blue module is roughly the SciPy community. The green module is the Wikimedia community. The purple module is the OpenStreetMap community.*

Sebastian Benthall (2015) “Testing Generative Models of Online Collaboration with BigBang.” Proceedings of the 14th Python in Science Conference, 175 - 181, 2015, Eds.Kathryn Huff and James Bergstra, 10.25080/Majora-7b98e3ed-01b

History

- 2015 - Developed to study open collaborative communities.
- 2016 - adapted to study human rights advocacy in IETF and ICANN
- 2020 - Article 19 funds improvements to gender and affiliation detection, IETF datatracker and attendance ingest.
- 2021 - Article 19 sponsors BigBang Sprint at IETF 110.
- 2021 - BigBang awarded funding from Prototype Fund



UNIVERSITY OF AMSTERDAM

Individual vs. Organizations in IETF

“Participation in the IETF or of its WGs is not fee-based or organizationally defined, but is based upon self-identification and active participation by individuals.” - Tao of IETF

Are the participants in IETF acting as individuals, or as parts of organizations (like companies?)

Normative questions, like:

- Are individuals better stewards of the public interest than commercial organizations?

A related, *descriptive*, question:

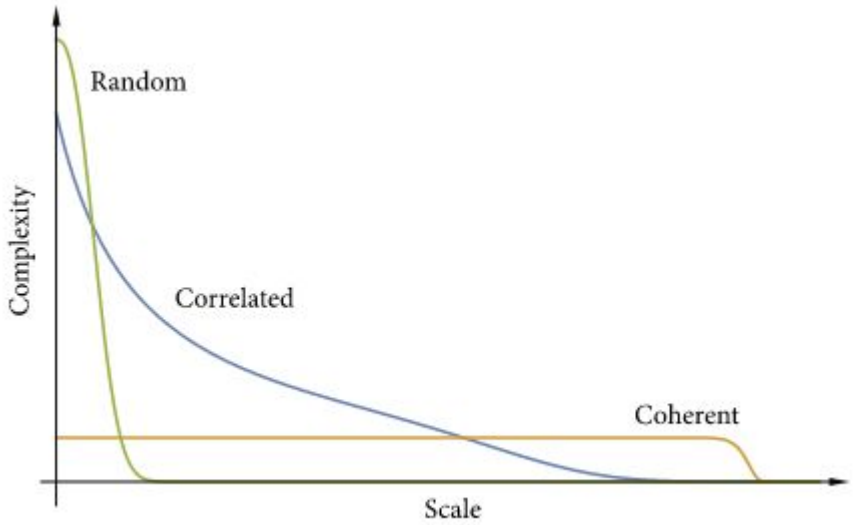
- How to determine when individuals are acting independently vs. as part of an organizational action.

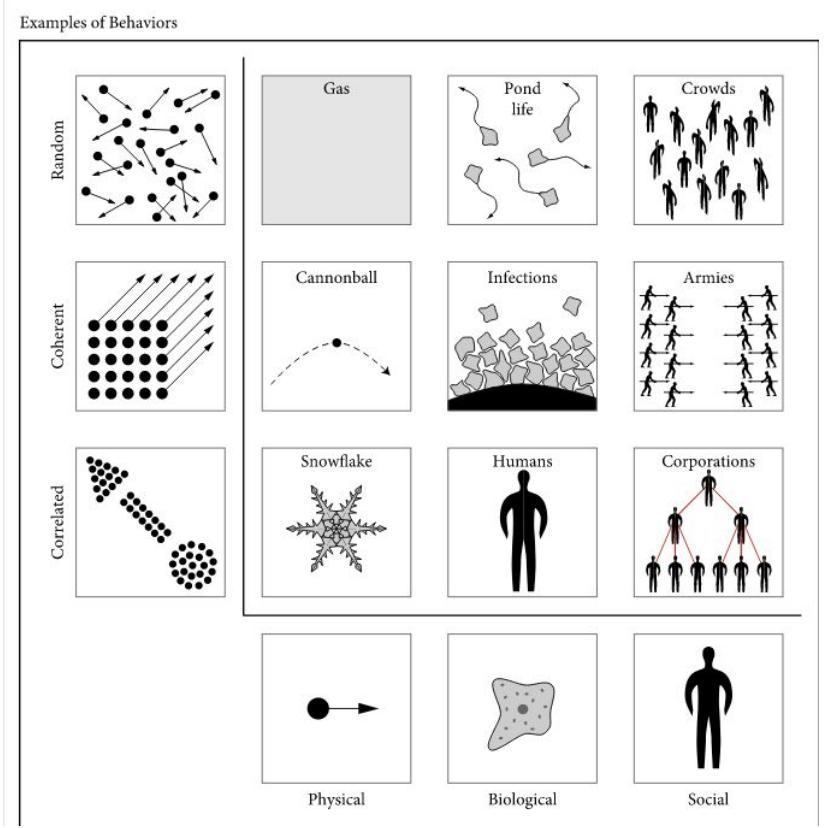
This is about how to do empirical work that spans *levels of abstraction*.

Tools, data, and methods

- Using BigBang for mailing list analysis
  - Getting participation in discussion outside of drafting
- This data is organized along multiple levels of abstraction.

Email Addresses		Working Group 1	Working Group 2
prefix_a	@domain_x.com	250	10
prefix_b		1	50
prefix_c	@domain_y.org	150	20
prefix_d		100	30

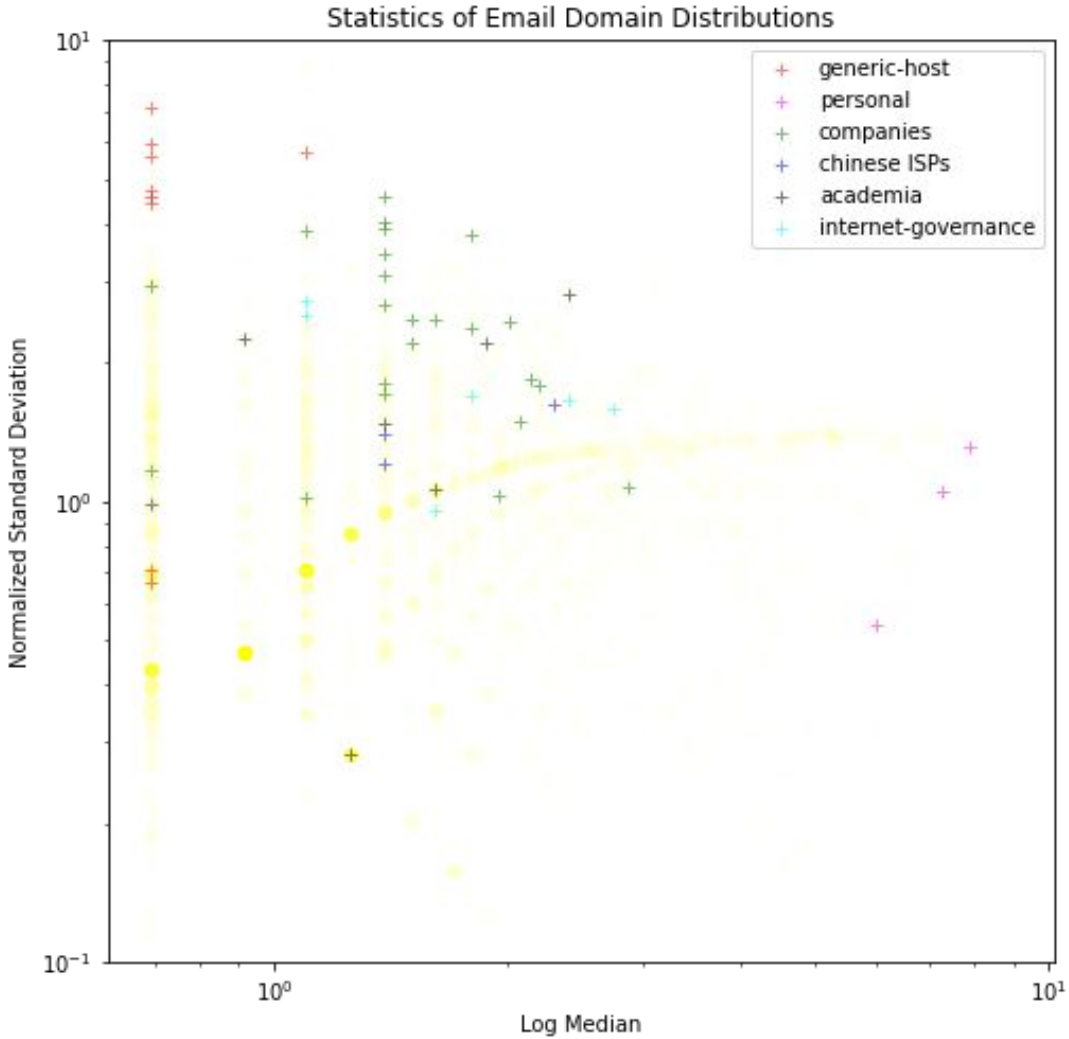




Use the *complexity profile* of a phenomenon to determine if it is acting randomly or else with a higher organizing principle. (Figures from Siegenfeld and Bar-Yam, *Complexity*, 2020)

Preliminary results on distributions over prefixes:

- *Generic email domains:*
  - e.g. `gmail.com`
  - *high standard deviation*
  - *low median*
  - R*andom* organization.
- *Organizational* email domains
  - E.g. `apple.com`
  - *higher median*
  - C*orrelated* organization
- *Personal* email addresses
  - E.g. `csperkins.org`
  - *low standard deviation*
  - *high median*
  - Coherent organization



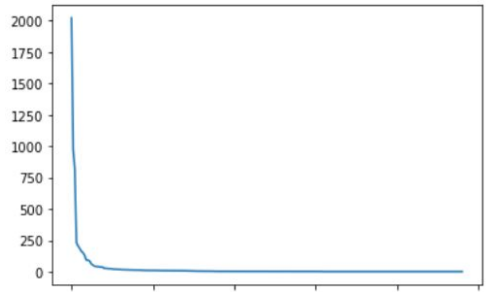
Next steps

- Consider organization within working groups
  - At individual level
  - At domain level
- Are the working groups random, correlated, or coherent organizations?
- Are they a mixture of activities of different types of organizations?

Questions and feedback: spb413@nyu.edu. Thanks!

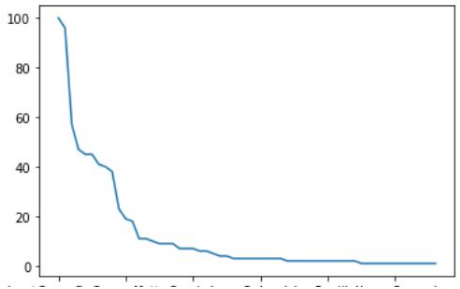
Messages to `httpbisa`

Messages per email - gmail.com



Many differently affiliated individuals at major differences in scale - random.

Messages per email - google.com



Area under curve indicating corporate strategy.