University of Glasgow

Internet Protocols Laboratory

An Empirical Analysis of the Internet Engineering Task Force with Computational Methods

Colin Perkins

Joint work with Mladen Karan, Prashant Khare, Ravi Shekhar, Matthew Russell Barnes, Waleed Iqbal, Junaid Qadir, Stephen McQuistin, Richard G. Clegg, Patrick Healey, Matthew Purver, Ignacio Castro, and Gareth Tyson

Queen Mary University of London

University of St Andrews

香港科技大學(廣州) THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Institut "Jožef Stefan" Ljubljana, Slovenija

Funded by the UK Engineering and Physical Sciences Research Council under grants EP/X027309/1, EP/R04144X/1, EP/S036075/1, EP/S033564/1, and DTA studentships

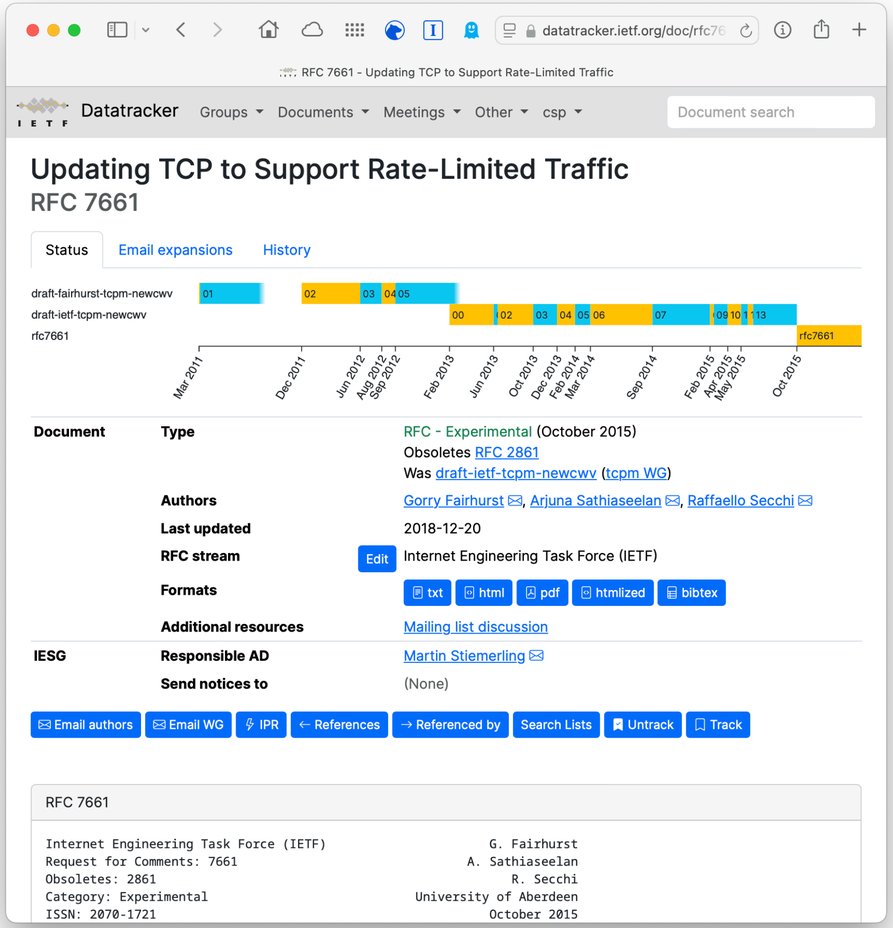
UKRI Engineering and Physical Sciences Research Council

Internet Engineering Task Force

- The IETF is the premier technical standards development organisation for the Internet
- Formed in the mid-1980s from the ARPANET project that created the precursor to the Internet
- Develops open, public, voluntary consensus standards – **RFCs** – that describe how the Internet works
  - TCP/IP, HTTP, email, WebRTC, TLS, BGP, …

IETF – Open Processes and Open Data

- IETF follows a policy of **aggressive openness**
  - Anyone may participate, no fixed membership
  - Email, teleconferences, in-person meetings (3x per year)
  - Makes public all RFCs, drafts, meeting recordings, minutes, presentations, review comments, approval ballots, patent declarations, participant lists, email discussion archives, …
  - Also available in machine readable form via a public API
- Unique dataset for studying collaborative online decision making, social dynamics, interpersonal communications, and development of Internet technologies



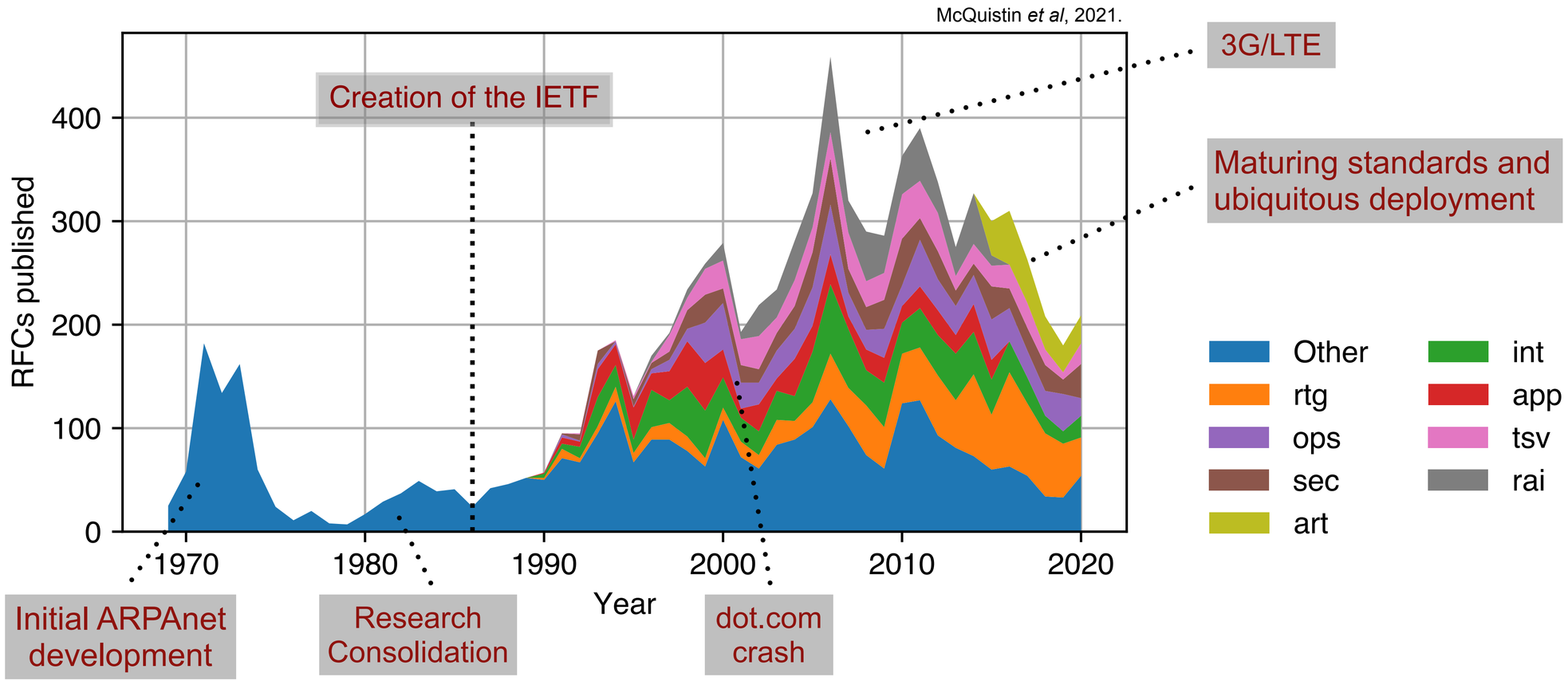
Goals of this Research

- **Enhance understanding of Internet standards**
  - Is the IETF effective at developing standards?
  - Who develops IETF standards?
  - Has the IETF transcended its US-centric origins to become a global standards organisation?
  - How do participants interact and communicate? Does the IETF show healthy organisational dynamics? Are those in leadership roles open to input from the wider community?
- **Enhance understanding of online decision making**
  - Improve understanding of social network analysis and natural language processing
  - Develop techniques to model decision making in a large online community
  - Informed by domain knowledge relating to IETF standards, Internet governance

Methodology

- Download from IETF datatracker, mail archive, and RFC index:
  - Metadata on 38,400 people, 140,000 documents
  - 2.5M emails, 75k addresses, 1,200 mailing lists
  - 8,711 RFCs from 6,200 authors
  - 6,759 RFC errata reports
- Perform entity resolution to find set of unique people and their affiliations
- Build social graph of email interactions
  - Labelled with dates, participant roles, documents mentioned, working groups; centrality (influence) and connectedness metrics
  - Linguistic analysis of communication patterns
- Based on this collected data, we studied:
  - RFC publication, complexity, and correctness trends over time
  - Trends in demographics and participant affiliation
  - Interaction between participants, trends in who is influential
  - Interaction style and use of language
  - Factors that affect success of documents and authors

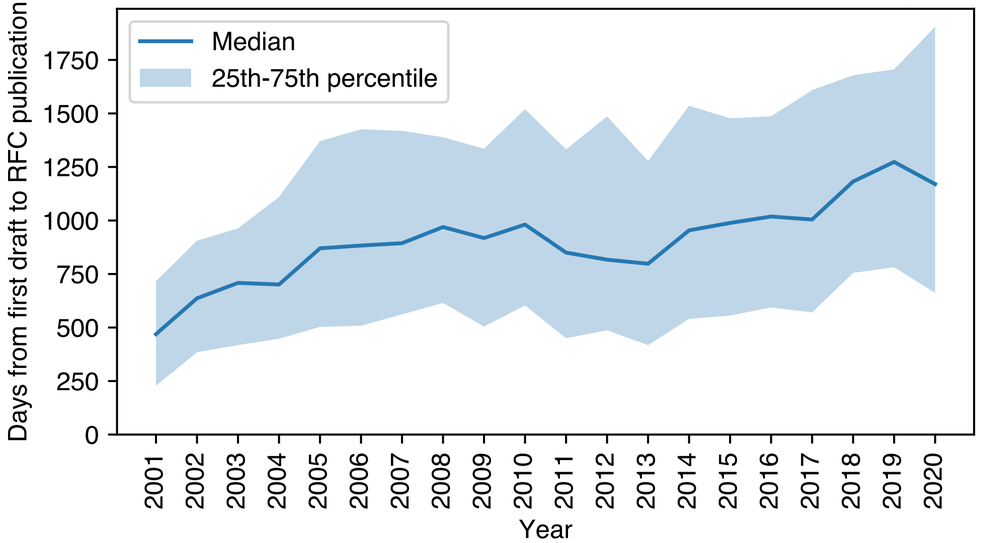
Trends in RFC Publication

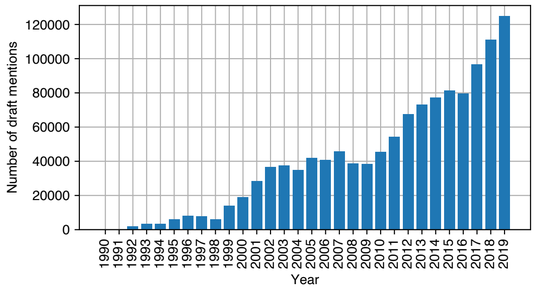


Complexity of Standards

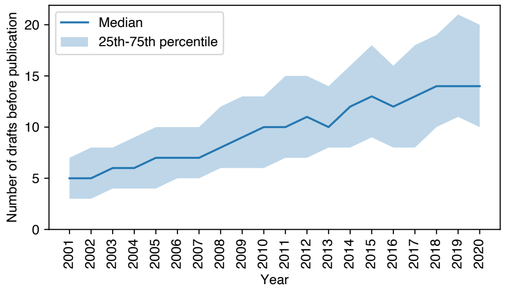
- Standards are taking longer to publish, but page counts remain broadly constant
  - The median number of days to publication was 469 in 2001, rising to 1170 in 2022
- The IETF is getting slower at publishing RFCs
  - Technical debt and increasing complexity?
  - Or natural progression in a maturing ecosystem?

McQuistin *et* al, 2021.

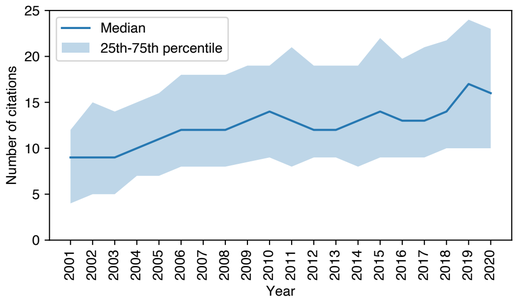




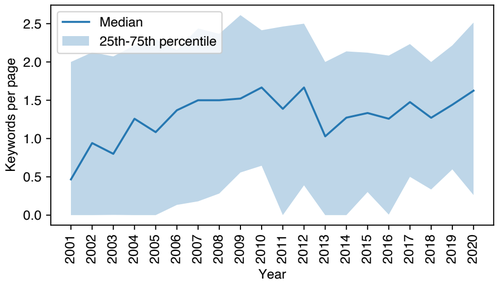
Number of emails mentioning drafts prior to publication



Median number of revisions made prior to publication has doubled

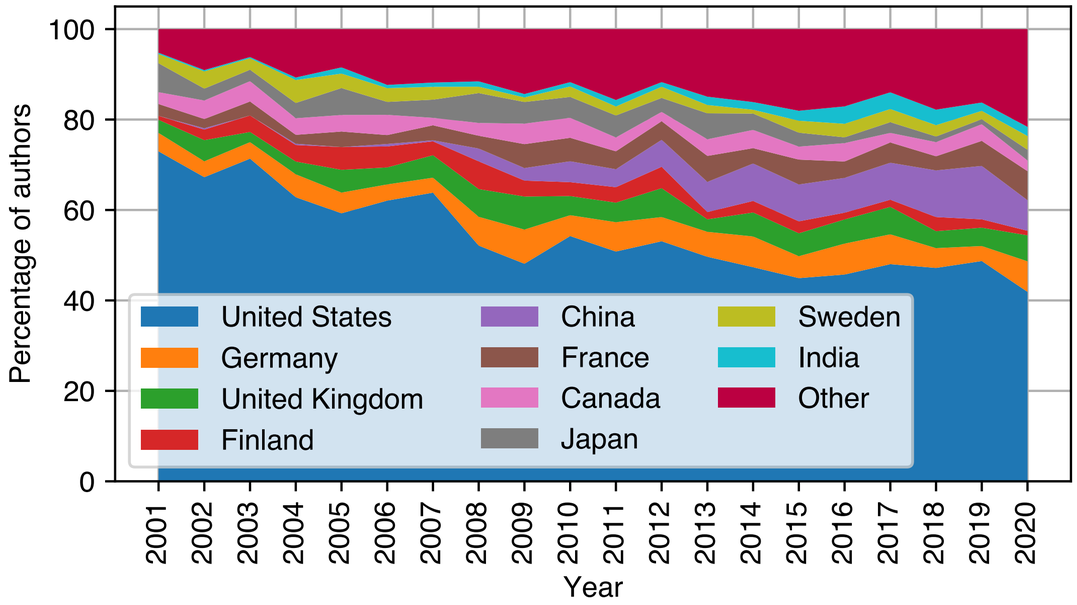


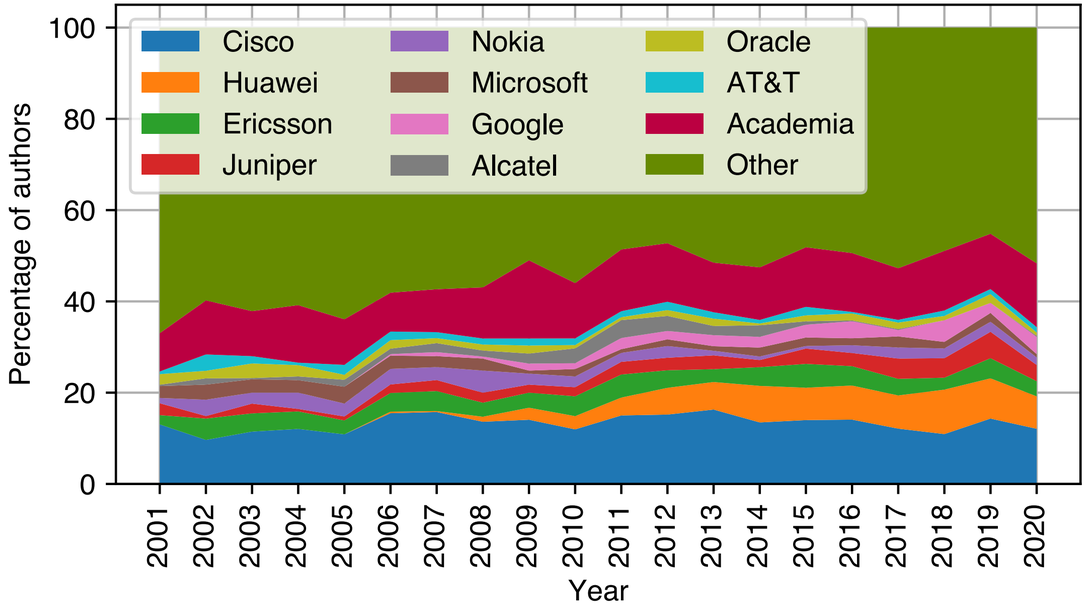
New drafts are citing increasing numbers of prior RFCs



Drafts are increasingly using normative language

Demographic and Affiliation Shifts

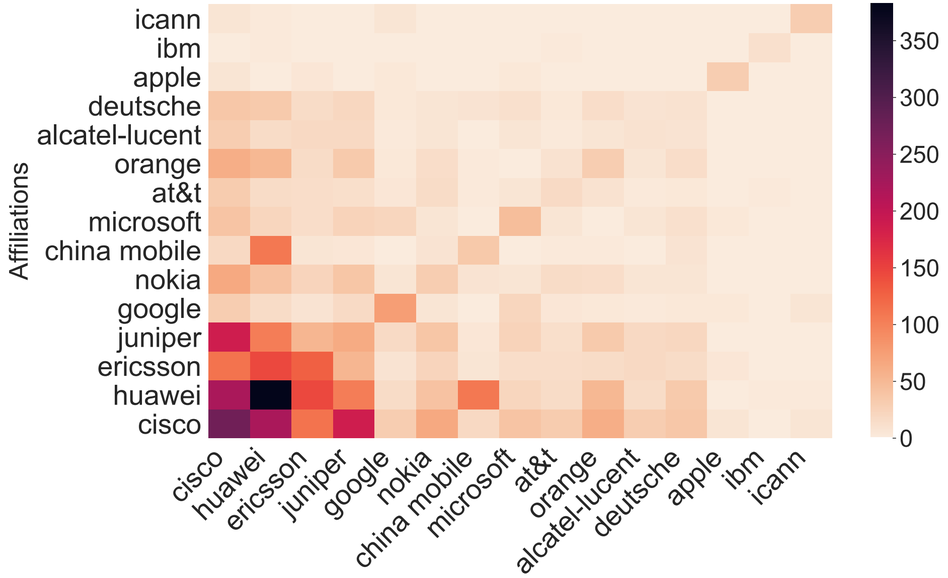




- Participation is increasingly multinational – shift towards Europe, China
- Strong tech company presence, but also academia, civil society, governments

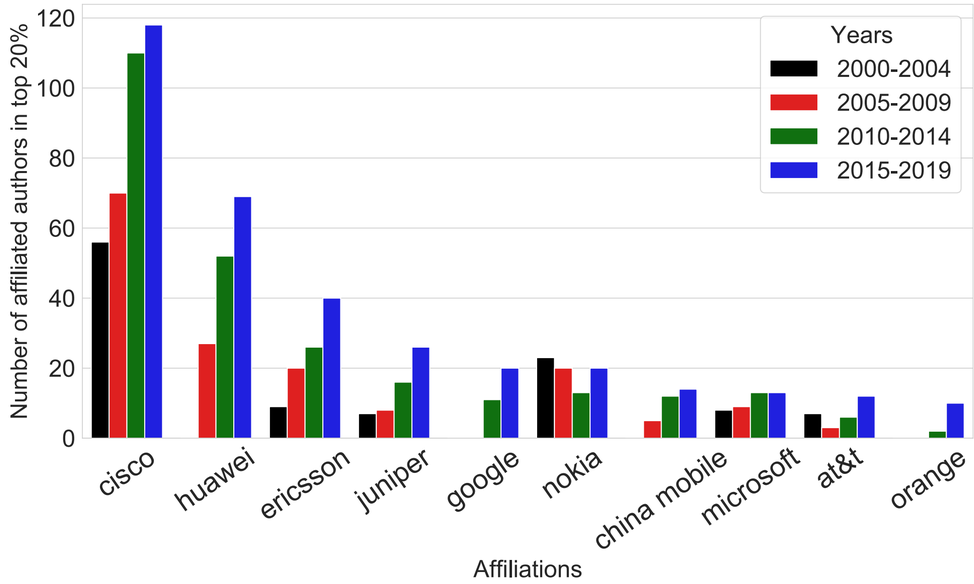
IETF Participation and Diversity

Khare *at al*, 2022



Authors mostly write with colleagues from same company, but strong cross-company collaborations (e.g., strong Cisco-Huawei co-authorship)

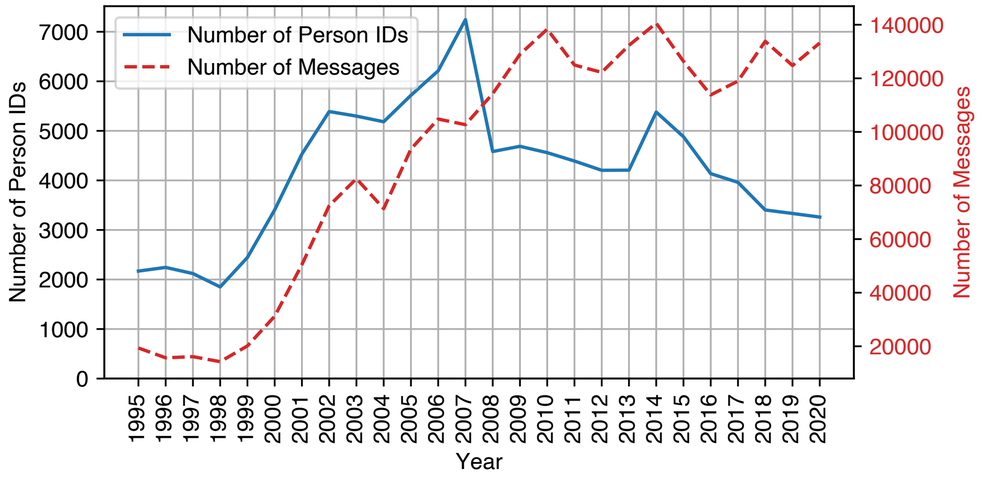
Khare *at al*, 2022



A small number of organisations employ an increasing fraction of prolific authors

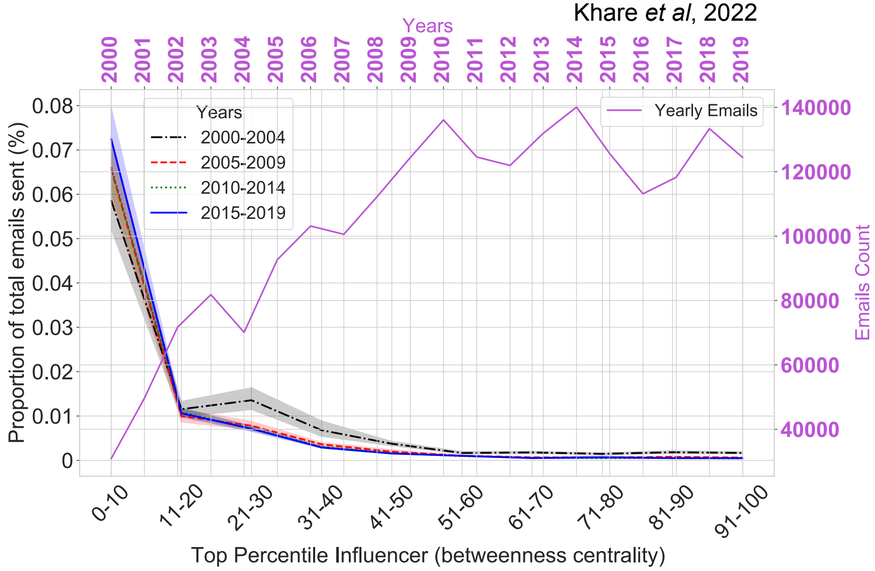
Social Graph – Communication Patterns

McQuistin *et* al, 2021.



Number of people involved peaked in 2006, when the number of RFCs published peaked, but number of emails sent has not declined

Khare *et al*, 2022



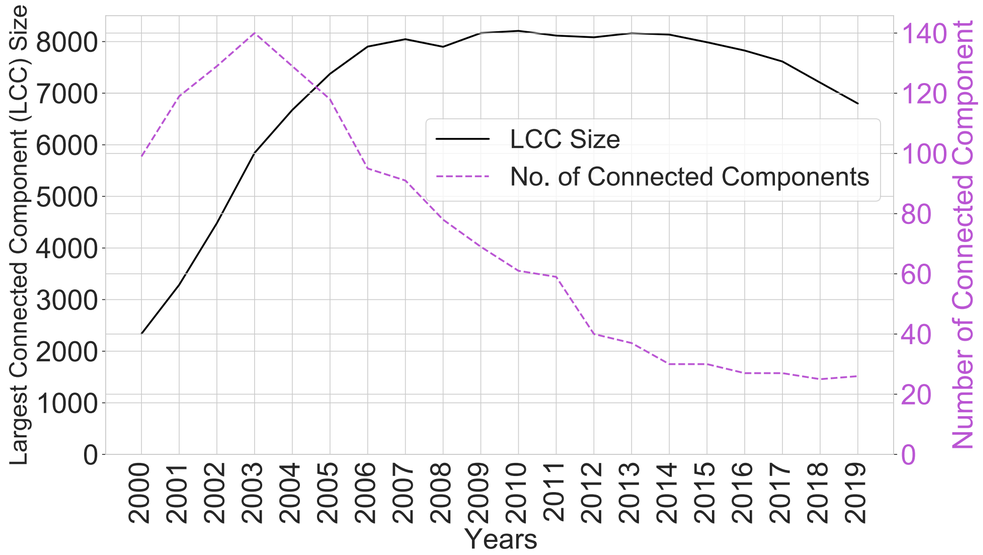
The most connected 10% of participants, by betweenness centrality, send 60% of emails

Communication overheads are increasingly a concern – how to make the process more efficient?

IETF appears strongly dependent on a small number of influential participants

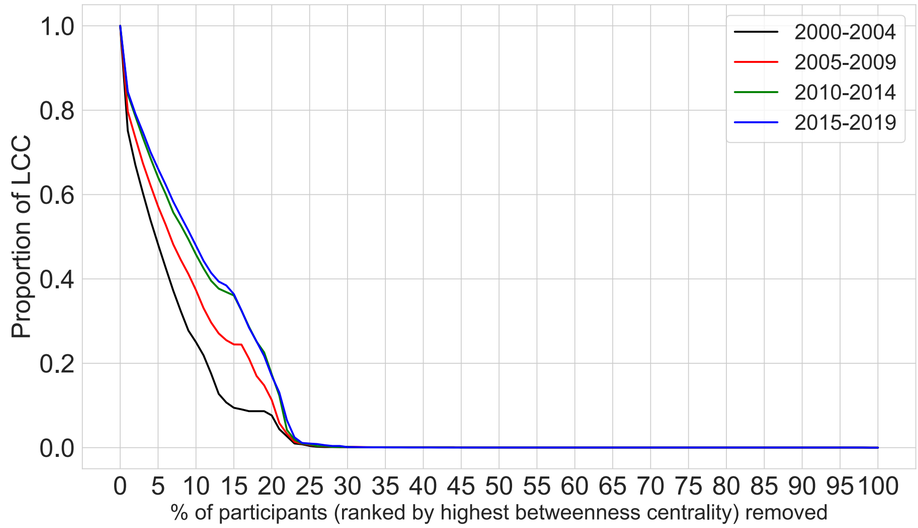
Social Graph – Communication Patterns

Khare *et al*, 2022



Fewer separate components of the email communication graph; largest connected component (LCC) is growing

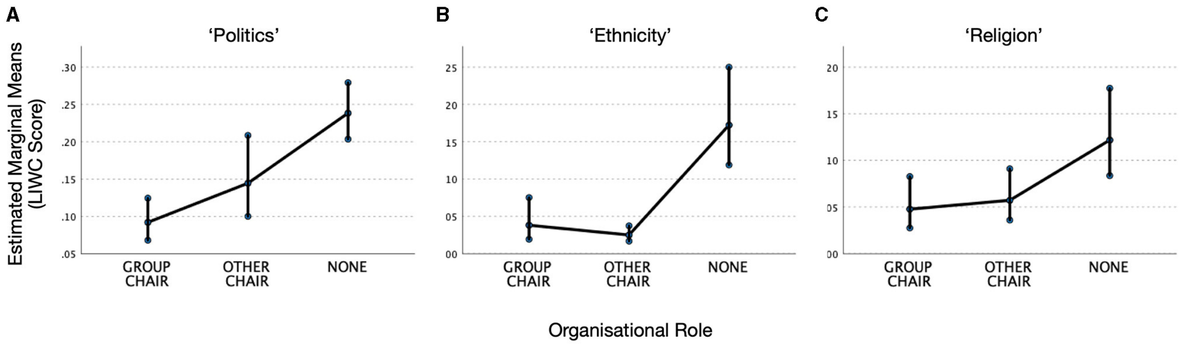
Khare *et al*, 2022



Removing the top 10% most connected people reduced size of LCC to 30% of original in 2000 but resilience has improved over time

IETF is becoming more cohesive – a small group of well-connected individuals still dominates, but the community as a whole is becoming better connected

Language and Communication



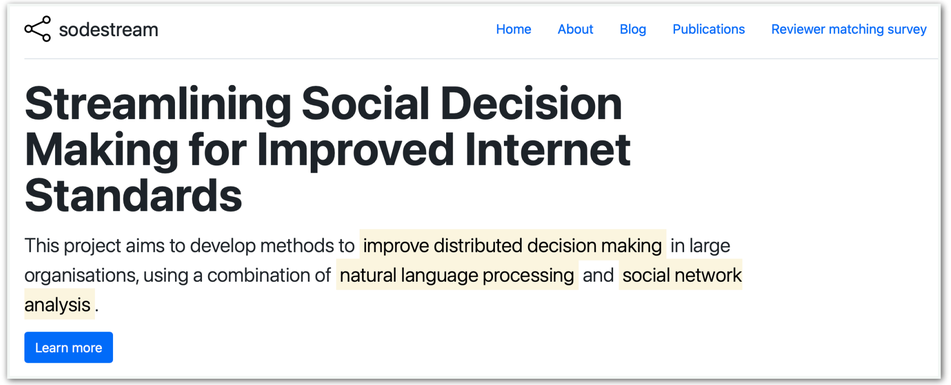
Language and influence	High influence	BIO, WE, INFORMAL, THEY, NEGEMO, ANGER, RISK, ADJECTIVE
	Low influence	SEXUAL, DEATH, INGEST, NETSPEAK, HEALTH, FEMALE, BODY, AFFILIATION, CONJ
Language and role	WG Chair influence	TENTAT, IPRON, SOCIAL, SEE, FEEL, WE
	non-WG Chair	COGPROC, RELATIV, AFFILIATION, I, REWARD
Changes in Language	Top 10 percentile	ADVERB, PREP, ANGER, AUXVERB, MALE, COGPROC, ACHIEV, RISK, FOCUSPRESENT
	Below 50^{th} percentile	FUNCTION, PPRON, SHEHE, IPRON, NUMBER, CERTAIN, SEXUAL, INFORMAL

Table 7: LIWC categories where $p < 0.05$.

- Use of language reflects organisational hierarchy – how people communicate shifts as they assume leadership roles
- Working group chairs less likely to use sensitive language, more collaborative and social
- Well connected people in the social graph are less formal, more social, but can also be more forceful
  - c.f., Cath: "Loud men, talking loudly"

Conclusions

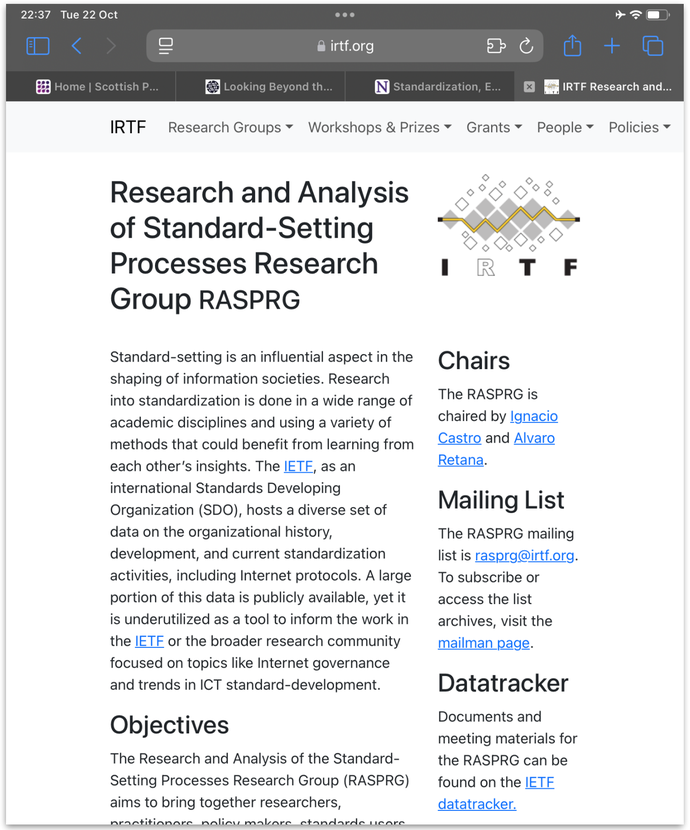
- IETF standards are essential for the operation of the Internet – but the itself IETF not well studied understood and differs significantly from some other SDOs
- Data reveals complex community dynamics, shifts in company influence and demographics, as the community grows away from its highly US-centric roots
- Our ongoing work considers community resilience and cross-SDO interactions
- Next directions:
  - Study the impact of non-commercial actors and consultants
  - Impact of patents on the IETF process



https://sodestream.github.io/

Advertisement: IRTF RASPRG

- The Internet **Research** Task Force helps make connections between researchers and the IETF standards community
- **Research and Analysis of Standard-setting Processes Research Group** aims to connect those studying Internet standards processes with those developing the standards
- Keen to make connections – to both help improve the way IETF works and to understand the Internet standardisation ecosystem more broadly



https://www.irtf.org/rasprg.html

References

1. Stephen McQuistin, Mladen Karan, Prashant Khare, Colin Perkins, Gareth Tyson, Matthew Purver, Patrick G. T. Healey, Walid Iqbal, Junaid Qadir, Ignacio Castro, “**Characterising the IETF Through the Lens of RFC Deployment**”, ACM Internet Measurement Conference, November 2021. https://sodestream.github.io/publications/characterising-ietf-rfc-deployment.pdf

2. Prashant Khare, Mladen Karan, Stephen McQuistin, Colin Perkins, Gareth Tyson, Matthew Purver, Patrick G. T. Healey, and Ignacio Castro, “**The Web We Weave: Untangling the Social Graph of the IETF**”, AAAI International Conference on Web and Social Media, June 2022. https://sodestream.github.io/publications/web-we-weave-icwsm.pdf

3. Stephen McQuistin, Mladen Karan, Prashant Khare, Colin Perkins, Matthew Purver, Patrick Healey, Ignacio Castro, and Gareth Tyson, “**Errare humanum est: What do RFC Errata say about Internet Standards?**”, Proceedings of the IFIP Traffic Measurement and Analysis Conference, June 2023. https://sodestream.github.io/publications/mcquistin2023errare.pdf

4. Mladen Karan, Prashant Khare, Ravi Shekhar, Stephen McQuistin, Ignacio Castro, Gareth Tyson, Colin Perkins, Patrick Healey, and Matthew Purver, “**LEDA: a Large-Organization Email-Based Decision-Dialogue-Act Analysis Dataset**”, Finding of the Association for Computational Linguistics, July 2023. https://sodestream.github.io/publications/karan2023leda.pdf

5. Prashant Khare, Ravi Shekhar, Mladen Karan, Stephen McQuistin, Colin Perkins, Ignacio Castro, Gareth Tyson, Patrick G. T. Healey, and Matthew Purver, “**Tracing Linguistic Markers of Influence in a Large Online Organisation”**, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, July 2023. https://sodestream.github.io/publications/khare2023tracing.pdf

6. Patrick G. T. Healey, Prashant Khare, Ignacio Castro, Gareth Tyson, Mladen Karan, Ravi Shekhar, Stephen McQuistin, Colin Perkins, and Matthew Purver, “**Power and vulnerability: managing sensitive language in organizational communication**”, Frontiers in Psychology, February 2024 https://dx.doi.org/10.3389/fpsyg.2023.1266425

7. Matthew Russell Barnes, Mladen Karan, Stephen McQuistin, Colin Perkins, Gareth Tyson, Matthew Purver, Ignacio Castro, and Richard G. Clegg, “**Temporal Network Analysis of Email Communication Patterns in a Long Standing Hierarchy”**, Proceedings of the International AAAI Conference on Web and Social Media, June 2024. https://arxiv.org/pdf/2311.13442.pdf