

Analyzing IETF Data (AID) Workshop



Internet Architecture Board

November 29 - December 2 2021



UNIVERSITY OF AMSTERDAM



IN-SIGHT:it

Citizenship and
standard-setting in
digital networks



Programme Committee: Niels ten Oever (chair, University of Amsterdam), Colin Perkins (chair, IRTF, University of Glasgow), Corinne Cath (chair, Oxford Internet Institute), Mirja Kühlewind (IAB, Ericsson), Zhenbin Li (IAB, Huawei), Wes Hardaker (IAB, USC/ISI).

There is a wealth of IETF data, how can it be used to:

- Improve the IETF and its processes ?
 - Create better standards that see more implementation ?
 - Aid IETF leadership, authors, and community members ?
 - Increase IETF legitimacy ?
-
- How to create a community around people that *use* IETF data ?



#	From		To
7	(For LLC services) A culture where people line up next to people with the same opinions and standing opposite those who disagree	→	(For LLC services) A culture where people line up together facing the data and working together to interpret it
8	(For LLC services) A set of long-standing issues left in the “too hard” basket as the culture of opinion and debate cannot resolve them	→	(For LLC services) An evidence led process in place to address all long-standing issues in the “too hard” basket.
9	Insufficient data on participants and their activities, motivations and expectations to support an evidence-led culture	→	Sufficient high-quality data on participants their activities, motivations, expectations, etc to support an evidence-led culture

From IETF Administrative Strategic Plan 2020

https://www.ietf.org/media/documents/IETF_Administrative_Strategic_Plan_2020.pdf

Workshop agenda

Monday: workshop & discussions

Tuesday: hackathon

Wednesday: hackathon

Thursday: workshop & presentations

Agenda

All times are UTC.

Monday, Nov 29

- 14:00: Opening (Ten Oever)
- 14:15: Tools, data, and methods (Arkko)
- 15:00: Observations on affiliation and industry control (Cath)
- 15:45: Community and diversity (Hardaker)
- 16:30: Break
- 17:00: Publications, process, and decision-making (Perkins)
- 17:45: Taking stock: questions and hacking groups (Kühlewind)
- 18:30: Closing (Cath)
- 19:00: End of Day

Tuesday, November 30

- 14:00-18:00: Hackathon
- 16:00-16:30: Sync-up

Wednesday, December 1

- 14:00-18:00: Hackathon
- 16:00-16:30: Sync-up

Thursday, December 2

- 14:00: Opening (Li)
- 14:15: Environmental Sustainability – initial results (Perkins)
- 14:45: Break
- 15:15: Result presentation (Ten Oever)
- 17:30: Wrap-up (Cath and Kühlewind)
- 18:00: End of Day

Many Data Sources

- Mailinglist archives
 - RFC Index
 - Datatracker (user account information, meeting proceedings, and minutes)
 - <https://notes.ietf.org/iab-aid-datatracker-database-overview#> (!)
 - IETF.org webpage visitor states
 - Github and interop testings
 - Survey data
-
- Combining data with that of other SDOs might provide interesting perspectives

Tools

The [ietfdata library](#): datatracker, RFC index, emailarchive

The [BigBang project](#): email subscriptions, network analysis, discourse analysis, user account data from datatracker

The [RFC Prolog Database](#): RFC content and metadata

[AUTHORSTATS](#) and [IETF Document Statistics](#)

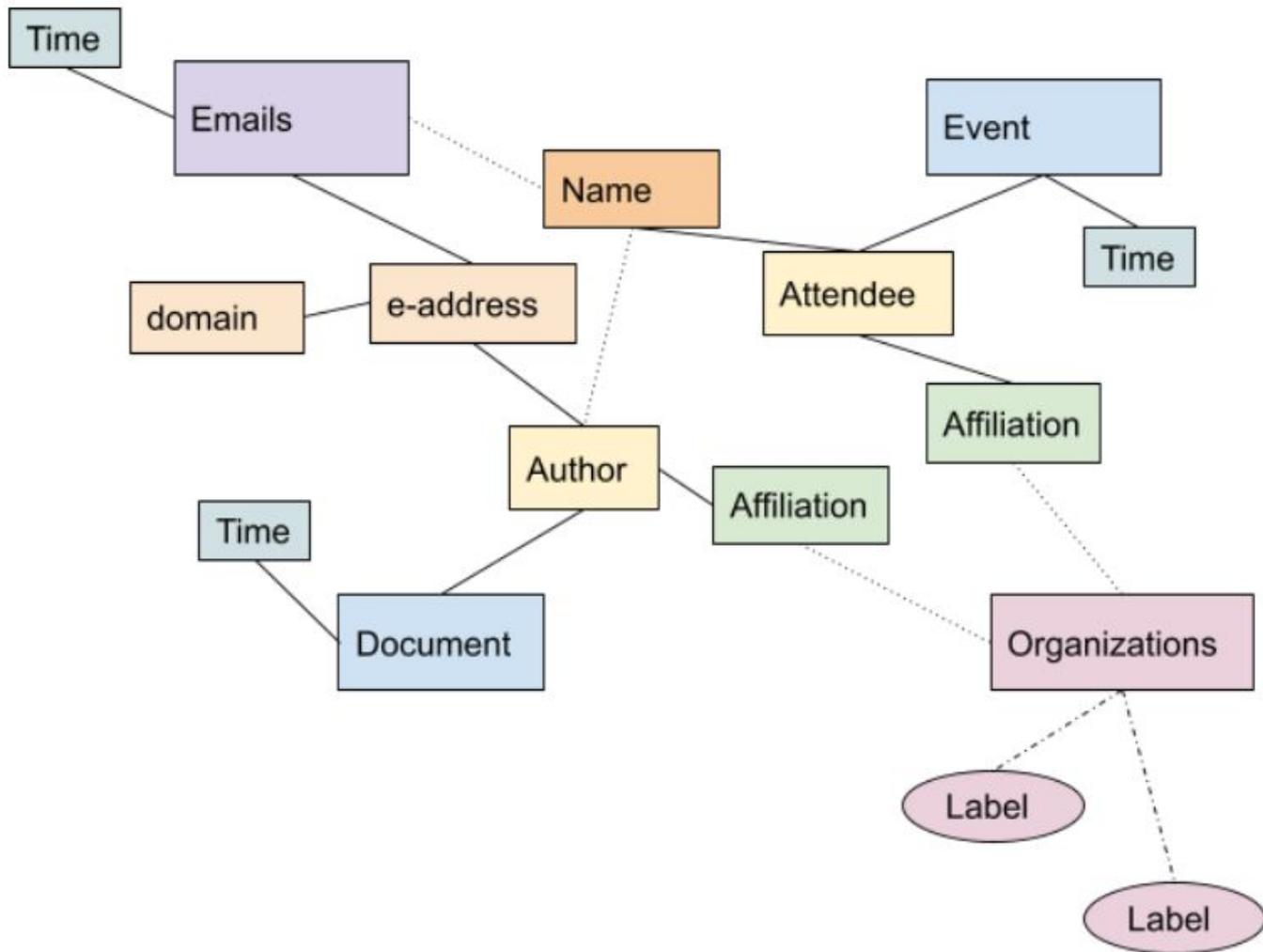
Observations on affiliation and industry control

Questions from papers:

- What are the relations between corporate interestes vs individual contributions\
- Is there competition for leadership positions?

Work done:

- Initial analysis on affiliation trends in mailinglist participation
- Develop database for stakeholder groups, categories, and mapping of subsidiaries
- Initial comparison between IETF and 3GPP data



name	category	subsidiary of / alias of	stakeholdergroup	nationality	email domain names	Membership Organization
International Institute of Information Technology - Bangalore	Research Institution		Academia	India		TSDSI
Internet Initiative Japan Inc.	Telecommunications Provider		Business	Japan	ij.ad.jp	
Internet Society	Internet Governance Body		Civil Society		isoc.org	
Internet Systems Consortium					isc.org	
Intersec SA	Software Provider, Consulting		Business	France		ETSI
Intertek	Testing and Certification		Business	United Kingdom		ETSI
IPCom GmbH & Co.KG	Technology research and development company		Business	Germany		ETSI
IPLOOK Networks Co. Ltd.	Networking Service Provider		Business	China		CCSA
IPV6 Forum	Association		Business	United Kingdom		OTHER
ISC	Internet Registry		Business	Civil Society		
Iskratel d.o.o. Ltd. Telecommunications Systems	Software Provider		Business	Slovenia		ETSI
ITOCHU Techno-Solutions Corporation	Software Provider		Business	Japan		TTC
ITRON SAS	Software Provider, Hardware Provider		Business	France		ETSI
Japan Radio Co. Ltd	Networking equipment vendor		Business	Japan		ARIB
Jetflow Technologies Co. Ltd	Technology research and development company		Business	China		CCSA
JMA Wireless	Networking Service Provider		Business	United States		ATIS
John Deere GmbH & Co. KG	Consumer hardware and software vendor		Business	Germany		ETSI
Johns Hopkins University Applied Physics Laboratory	Research Institution		Academia	United States	innovationslab.net	ATIS

```
In [91]: determine_sector('Q217082') # IETF
```

```
VERBOSE:pywiki:Found 1 wikidata:wikidata processes running, including this one.
```

```
Out[91]: 'technical'
```

```
In [94]: determine_sector('Q168756') # UC Berkeley
```

```
Out[94]: 'academic'
```

```
In [95]: determine_sector('Q376150') # ITU
```

```
Out[95]: 'intergovernmental'
```

```
In [96]: determine_sector('Q5059580') # CDT
```

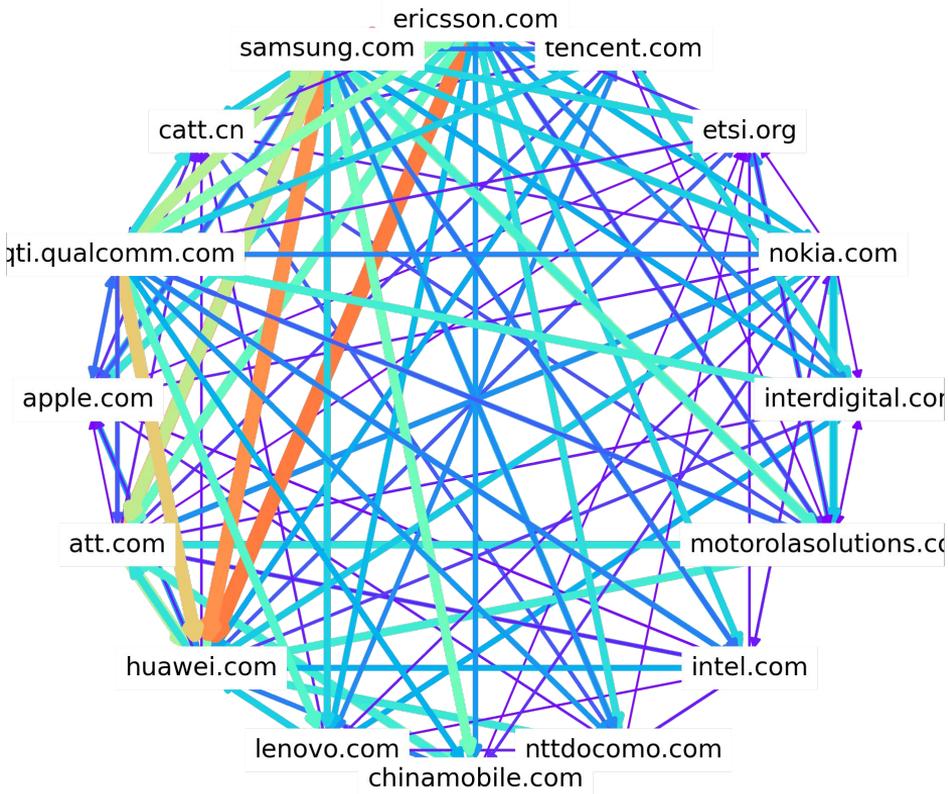
```
Out[96]: 'nonprofit'
```

```
In [97]: determine_sector('Q95') # Google
```

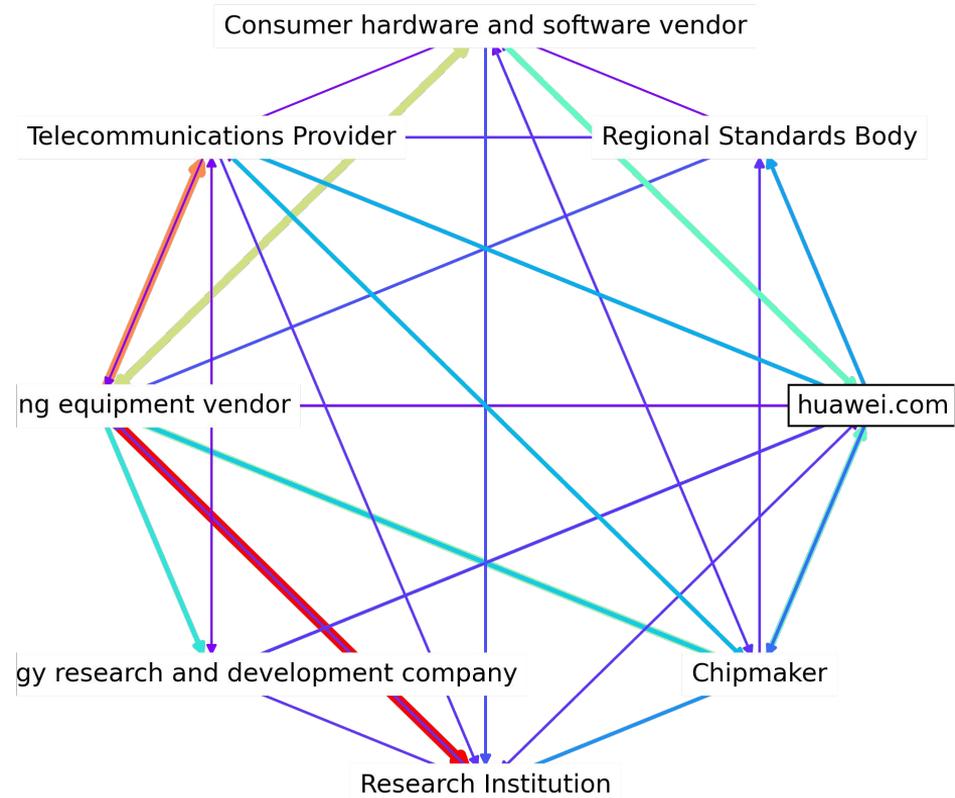
```
Out[97]: 'business'
```

	name	category	subsidiary of / alias of	stakeholdergroup	nationality	email domain names	Membership Organization	Wiki Page		norm_org	max_domain_jetf
18	Adobe	Software provider	NaN	Business	United States	adobe.com	NaN	NaN		Adobe	gbiv.com
23	Afilias	Internet Registry	NaN	Business	United States	NaN	NaN	NaN		Afilias	afilias.info
26	Akamai Technologies	Content Distribution Network	NaN	Business	United States	akamai.com	NaN	NaN	Akamai Technologies		akamai.com
29	Alibaba (China) Group, Ltd.	NaN	NaN	Business	China	alibaba-inc.com	CCSA	NaN		Alibaba	alibaba-inc.com
36	Amazon Web Services Inc.	Cloud Provider	NaN	Business	United States	NaN	ETSI	NaN	Amazon Web Services		amazon.com
38	AMS	IETF secretariat	NaN	Governance body	United States	NaN	NaN	NaN		AMS	amsl.com
44	APNIC	Internet Registry	NaN	Governance body	NaN	NaN	NaN	NaN		APNIC	apnic.net
45	Apple	Consumer hardware and software vendor	NaN	Business	United States	apple.com	NaN	NaN		Apple	apple.com
62	Arista Networks	Cloud Provider	NaN	Business	United States	arista.com	NaN	NaN	Arista Networks		arista.com
63	Arrcus	Cloud Provider	NaN	Business	United States	arrcus.com	NaN	NaN		Arrcus	arrcus.com
64	ARTICLE19	Civil Society Organization	NaN	Civil Society	United Kingdom	article19.org	ETSI	NaN		Article19	article19.org

3GPP_TSG_SA_WG6



3GPP_TSG_SA_WG6



Community and Diversity

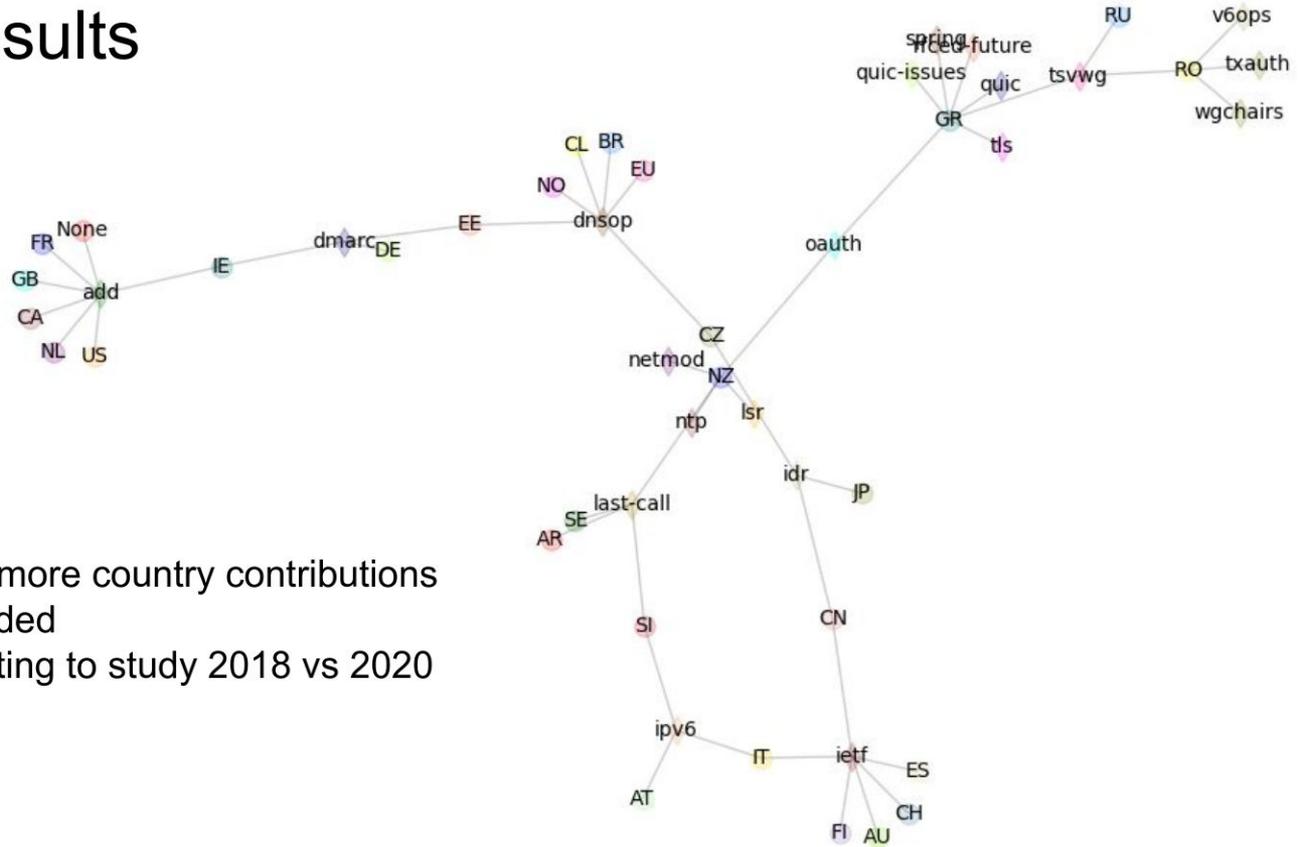
Questions from papers:

- Who is in the IETF, who is not there? What is the existing diversity? Can trends/reasons be found using data? What could be different if diversity had been stronger?

Discussion / Hacking

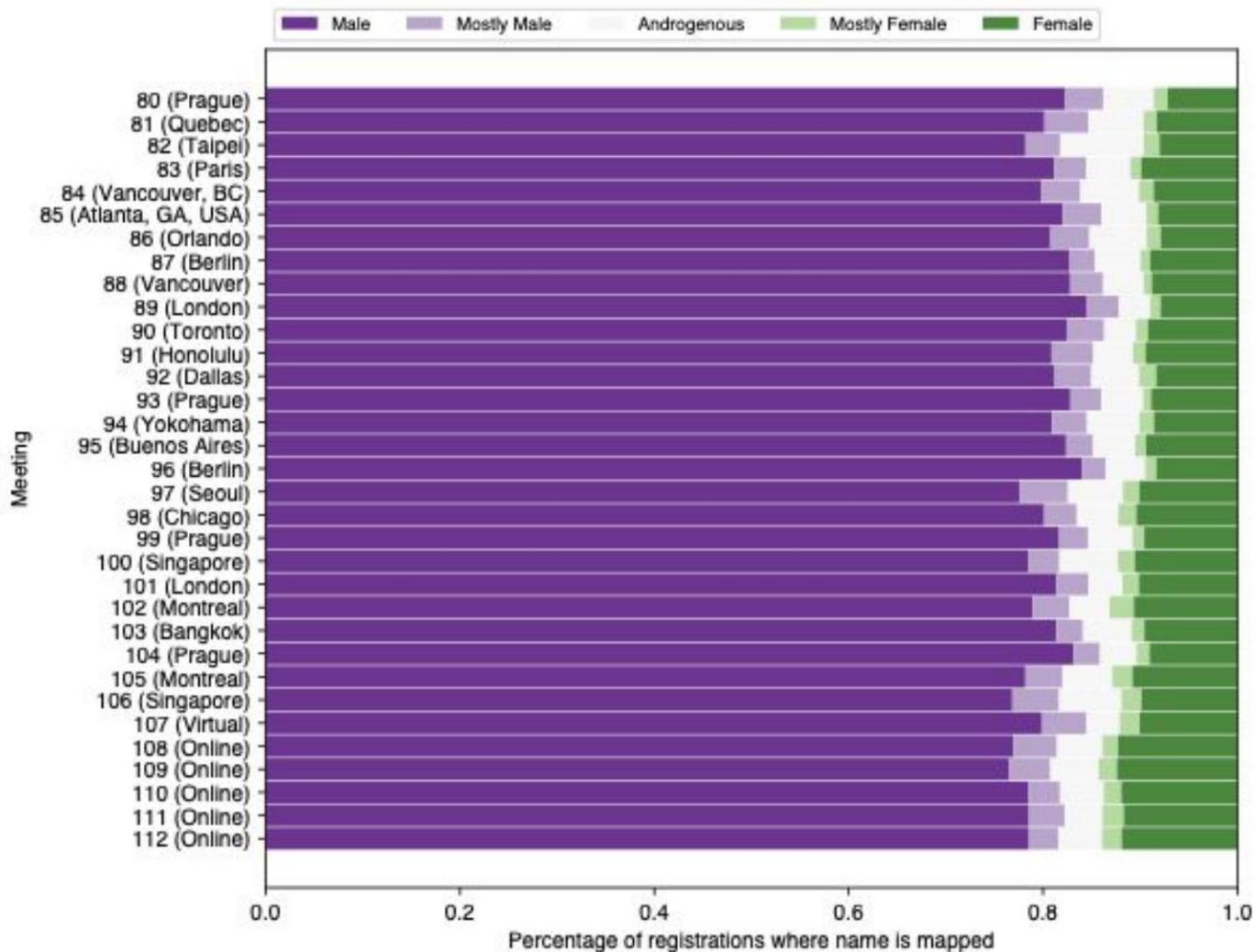
- What can be measured and what can not be (evaluation of tools and approaches)
- Studying of gender identification, country/region participation, and tenure
- Considerations on privacy and ethics

Very-initial results



Note:

1. There should be more country contributions
2. More data is needed
3. Would be interesting to study 2018 vs 2020



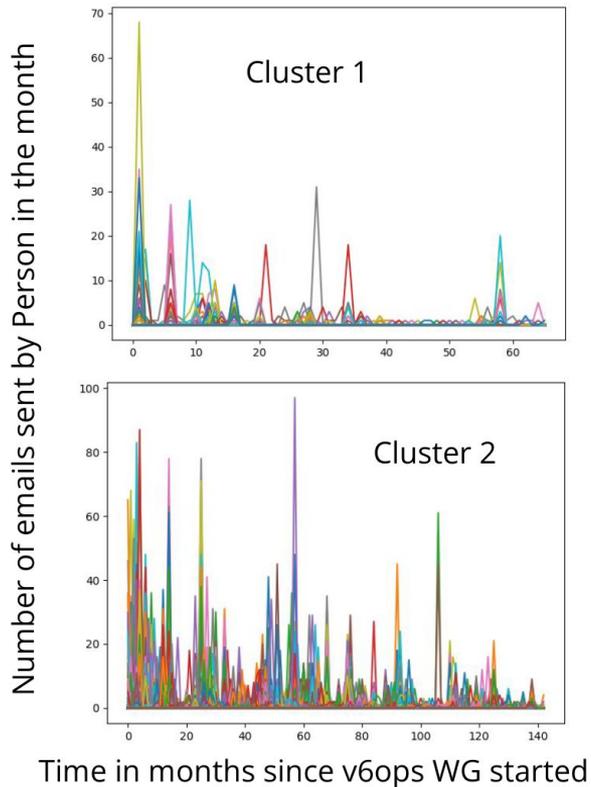
Publications, process, and decision-making

Questions from papers:

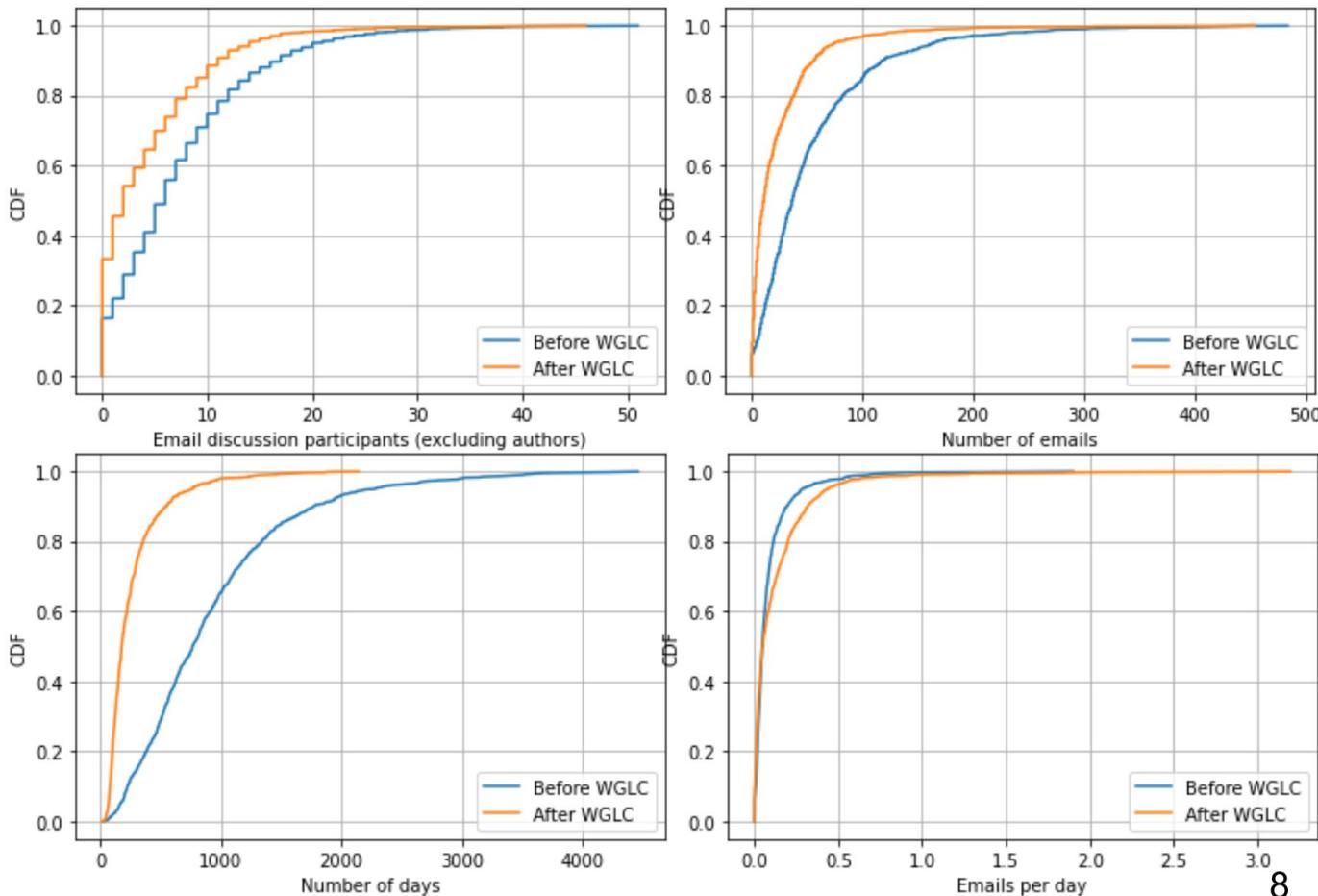
- How are decisions made in the IETF? Can these decisions be studied and can they improve decision making processes for authors, leadership, etc?
- What makes a successful RFC?

Discussion / Hacking

- Who writes RFC ? Newcomers vs oldtimers? Affiliation? Nationality?
- Data driven mining of homophilic users
- Requirements for NLP to understanding decision making process
- How to measure deployments
- Measuring efficient of RFC development process. Where are common delays?
- When does WGLC take place?



- First level grouping of homophilic participants by their temporal activity within v6ops working group in 10 clusters
 - People in Cluster 1 are interested in different aspects of v6ops than those in Cluster 2.
 - **Jari Arkko, Mirja Kuhlewind** in Cluster 1
 - **Timothy Baldwin, Michael Richardson, Paul A. Vixie** in Cluster 2
 - **Fred Baker** in another cluster
 - Activity peaks consistently overlap indicating they are interested in similar issues - help identify advocates? Leaders?
- **Entity Disambiguation** - Same name, different email address found in same cluster - eg Eric Klein, Linjian Song, Christopher Liljenstolpe, m7m7, etc



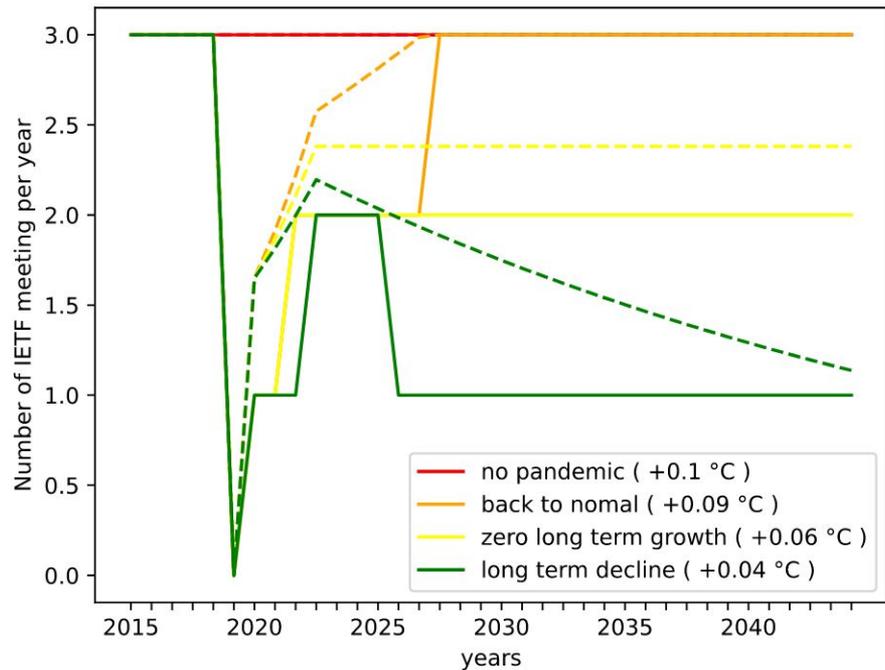
Environmental Sustainability

Questions from papers:

- How 'green' in the IETF
- How sustainable are IETF meetings?

Discussion / Hacking

- Are, can, and/or should sustainability considerations be part of the RFC development process
- How can CO2 emission per person per meeting be reduced?



Applying these scenarios to the IETF seems to indicate 1 IETF meeting per year

Observations

- Attendance in the workshop from IETF-ers, to social scientists and data scientists
- Mix of sessions + hackathon worked very well
 - **New collaborations**
 - **Code and toolchains tried and tested**
 - **New research questions developed**
- Basis for an research agenda for IETF data to inform future research and development
- Follow-up workshop under discussion



Other Links

Papers: <https://www.iab.org/activities/workshops/aid/>

Slides: <https://github.com/intarchboard/workshop-aid/tree/main/slides>

Draft Report: <https://intarchboard.github.io/workshop-aid/draft-iab-aid-workshop.html>

Video Recording: <https://www.youtube.com/watch?v=AAliAmBpzII&t=3s>