



**UiO**  **Faculty of Mathematics and Natural Sciences**  
University of Oslo



# **IAB AID workshop hackathon: Decision making and success factors for creating RFCs**

Michael Welzl



*IAB AID workshop*  
*30. 11. 2021*

# Some questions that were discussed

## Draft submission by newbies

- Sinha et al: Based on the analysis by [3] and [5, 6], it is evident that senior members are more active on the mailing lists and therefore may have a greater influence. It is expected then, that dialogue resulting in consensus for a newbie to the IETF would require advocacy from senior members irrespective of technical merit. There are well known long standing contentious issues that are debated regularly. It is not yet known if the delay in consensus is due to factions or lack of technical clarity. It is not also known, whether there are groups of members who are regularly opposed and consistently fail to obtain consensus despite technical merit.
- Griwodz et al: Does seniority improve the chances of success? But do they face less or more resistance in discussions, or is the atmosphere in discussions different from that faced by newcomers? Does the atmosphere in discussions for proposals that face active resistance develop over time, or does it stagnate?

# Some questions that were discussed /2

## Procedural steps towards RFC publication

- Sinha et al.: By studying the volume, diversity and consistency of engagement on mailing lists and meeting minutes historically versus recently, we can understand whether the virtual meetings has (..) lowered the impedance to seeking consensus.
- Griwodz et al.: Correlation of discussion with the development of drafts (email content); WGLC analysis

## Who did what...

- Priyanka Sinha and Amit Kumar worked on mailing list analysis
  - Identify leaders (influencers); some overlap with ongoing work of Colin Perkins et al
- Carsten Griwodz + Michael Welzl worked on WGLC analysis
  - Colin Perkins gave technical support 😊

# What worked and what didn't work

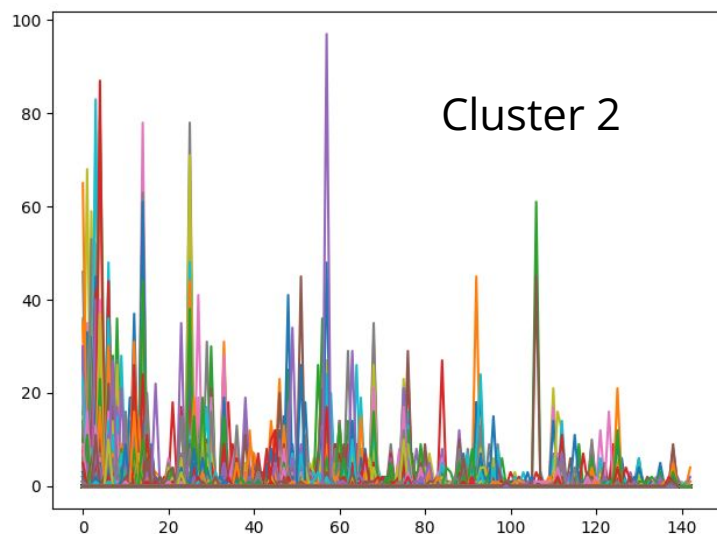
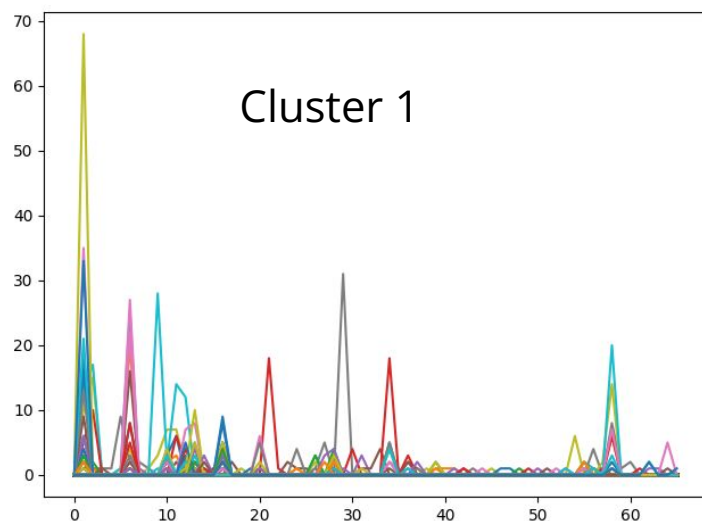
- What didn't: mongodb on colab (at first)
  - ietfdata on colab: python version problem
  - Needs python3.9 as the default python version; the python 3.7 pip version of ietfdata is the 0.4 version
- What worked:
  - ietfdata library: we're having a lot of fun with this
- Possible future work, next steps
  - Stay in touch for upcoming NLP work: Priyanka and Michael talked about sharing models
    - NLP models need to be fine-tuned for IETF "language"

# Some results

# IAB AID Workshop 2021 Hackathon Findings

- Google **colab** notebooks successfully ran **ietfdata** 0.4.0 library
  - python 3.7
  - MongoDB with persistent storage in Google Drive
  - Also installed ietfdata 0.5.3 library from source (github)
    - Requires python3.9 as default Python to run
- Email address to person entity extraction and disambiguation is hard
  - Downloaded **v6ops** working group mailing list >30,000 emails and adapted example script **uniq-emails.py** to obtain addrs.json (>700)
  - Several email address not resolved to persons (>400)
  - Several email addresses either have no full name or list of alternate names or none, such as Dave, organizations such as IETF Secretariat, etc
- Mined homophilic participants of v6ops WG - Priyanka Sinha, Lipika Dey, Pabitra Mitra, and Dilys Thomas. 2020. A Hierarchical Clustering Algorithm for Characterizing Social Media Users. In Companion Proceedings of the Web Conference 2020 (Taipei, Taiwan) (WWW '20). Association for Computing Machinery, New York, NY, USA, 353–362. <https://doi.org/10.1145/3366424.3383296> <https://www.youtube.com/watch?v= OmQWp7vflg>

Number of emails sent by Person in the month



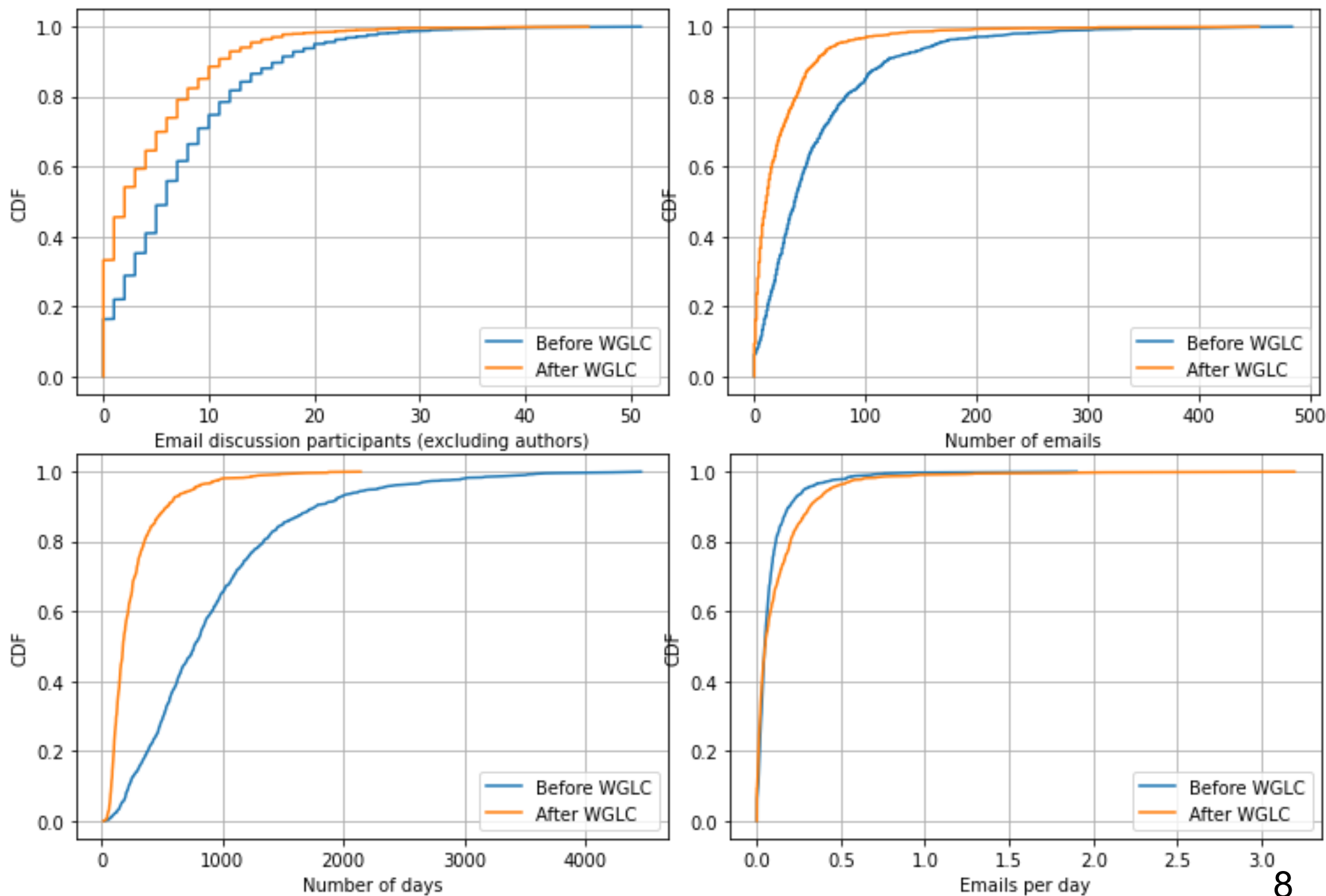
Time in months since v6ops WG started

- 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



# WGLC analysis using ietfdata

- Per RFC: WGLC number and dates available in datatracker
  - We analyzed the mailing list discussions related to the first WGLC
- For the set of precedings drafts of an RFC
  - Find: publication date range (first-WGLC, WGLC-RFC)
  - Emails related to this range: filter by date and subject, check for key part of draft name: draft-ietf-key-part-05
- Left this running today... 94 WGs, 1141 RFCs



# Thank you!

## Questions?