Group 2: Diversity/Inclusivity and general participants characteristics

IAB AID Hackathon

International contributions to working groups

Methodology:

- Take top 20 mailing lists from 2020
- Parse headers to examine source networks of contributions
- Collect in a node/edge graph
- Plot

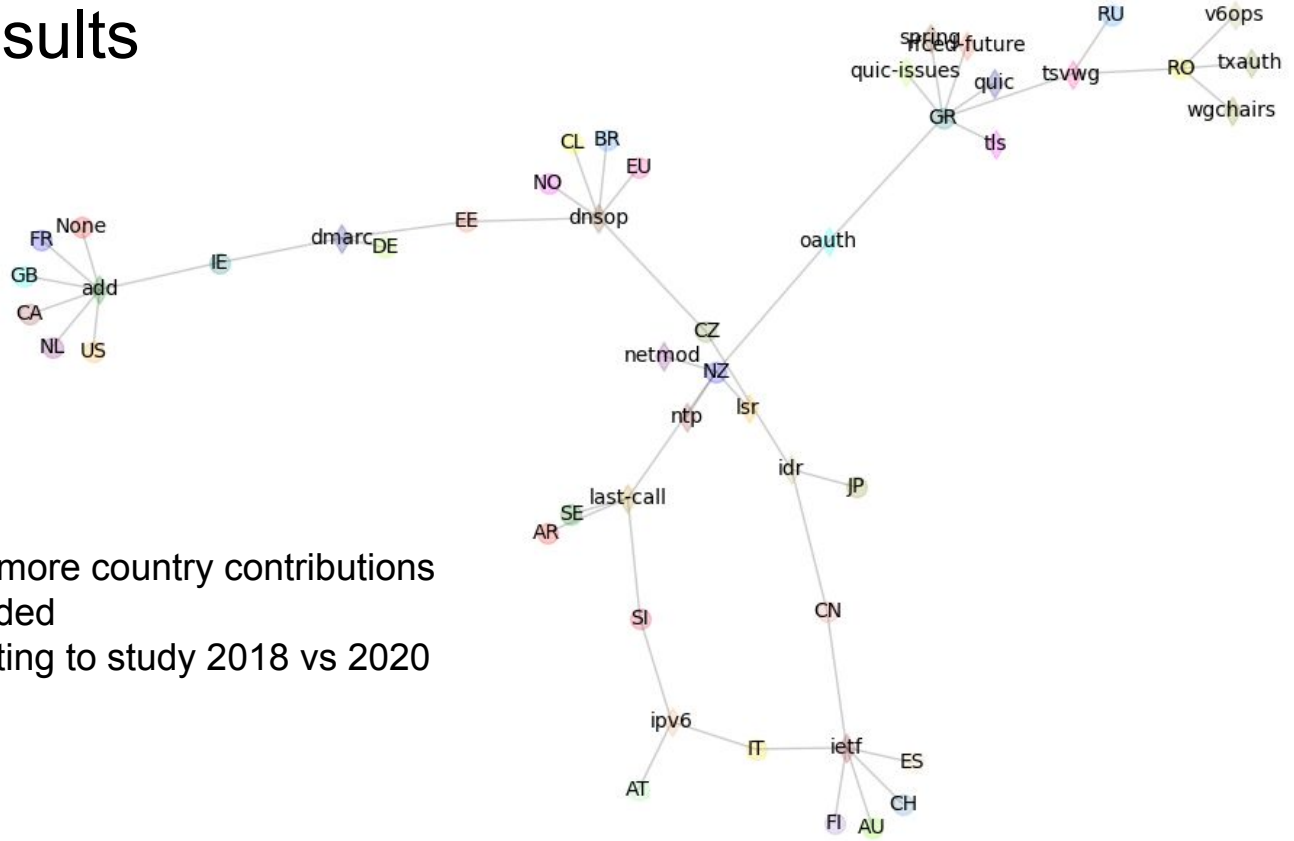
Issues:

- Determining source country is, of course, challenging

Results:

- Proof-of-concept success looks interesting (next slide)
- Future work is needed to pull in more data and clean better
- Future work is needed to use datatracker/etc for country derivations when possible
    - Outstanding question: do you believe the country they’re coming from or what is reported in data-tracker?

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

Authorstats tools (very limited hacking)

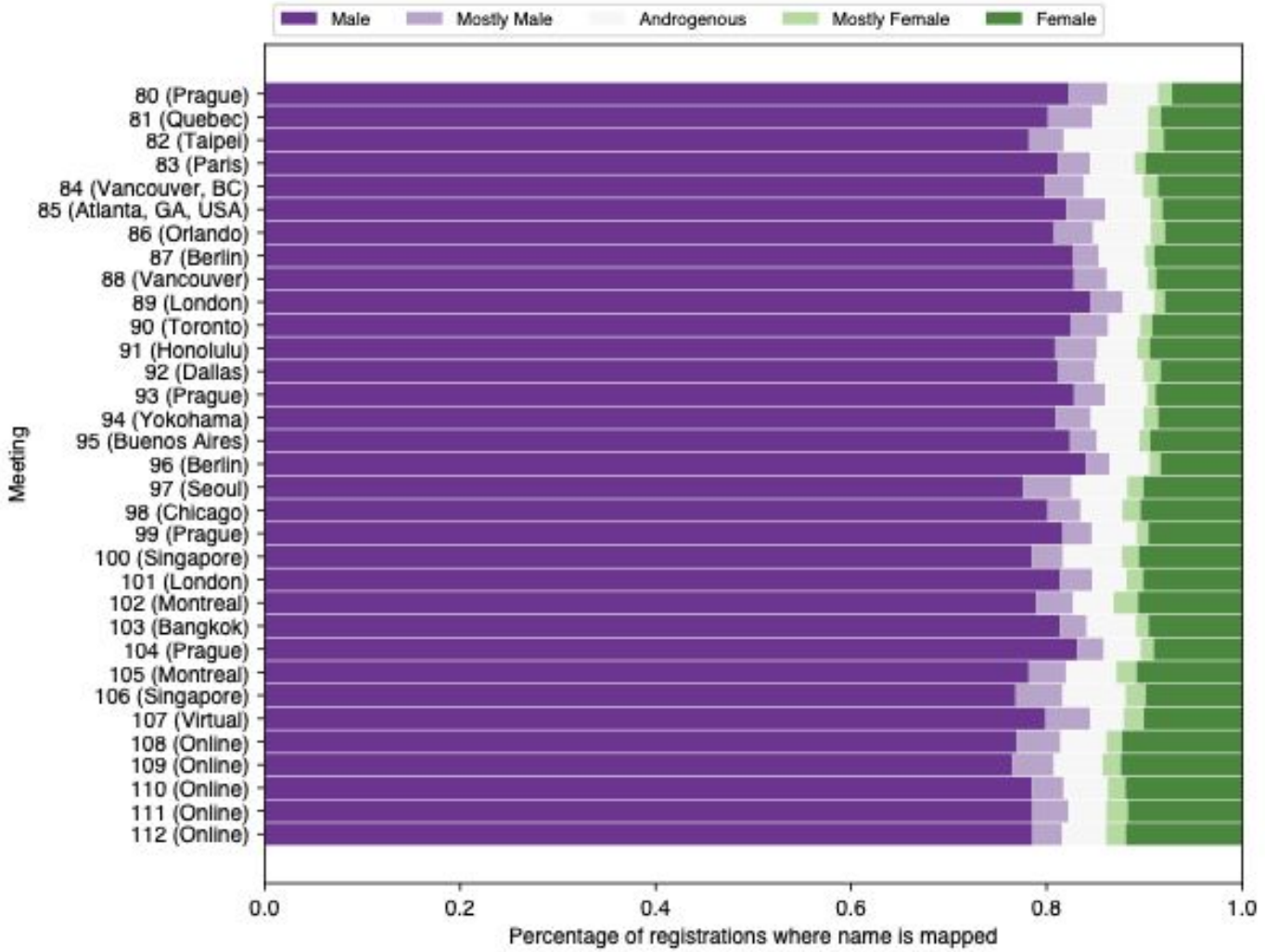
- Gender analysis is currently discontinued (could bring back aggregate, but needs work)
- Looked into using organization classifications from Niels, Nick, etc. -- if classifications into academic, operator, vendor, etc. were available, that data could be highly interesting, and provide some insight to the IETF process.
    - Did not have time to complete this during this week, but looking into it

Privacy Considerations of IETF Data Research

- IETF has been conservative about collecting any extra data (and ask for permission)
- Transparency of the process and public good are reasons for having some data and doing research on what's going on
- Nevertheless, important to follow ethical and legal principles
- Aggregation, avoiding displaying or collecting unnecessary information, etc.

Gender analysis of meeting registrations (Stephen)

- Methodology:
  - Fetch `first_name` field from meeting registrations from the last 11 years
  - Feed `first_name` into the Python `gender_guesser` library
- Notes
  - The library doesn’t map 20-25% of names
  - There are almost certainly biases in the dataset the library uses, but these aren’t clear
  - Would be better to have self-declared gender identity, as in the IETF survey (has started to be asked in 2021)


Male
Mostly Male
Androgenous
Mostly Female
Female
80 (Prague)
81 (Quebec)
82 (Taipei)
83 (Paris)
84 (Vancouver, BC)
85 (Atlanta, GA, USA)
86 (Orlando)
87 (Berlin)
88 (Vancouver)
89 (London)
90 (Toronto)
91 (Honolulu)
92 (Dallas)
93 (Prague)
94 (Yokohama)
95 (Buenos Aires)
96 (Berlin)
97 (Seoul)
98 (Chicago)
99 (Prague)
100 (Singapore)
101 (London)
102 (Montreal)
103 (Bangkok)
104 (Prague)
105 (Montreal)
106 (Singapore)
107 (Virtual)
108 (Online)
109 (Online)
110 (Online)
111 (Online)
112 (Online)
Meeting
0.0
0.2
0.4
0.6
0.8
1.0
Percentage of registrations where name is mapped