

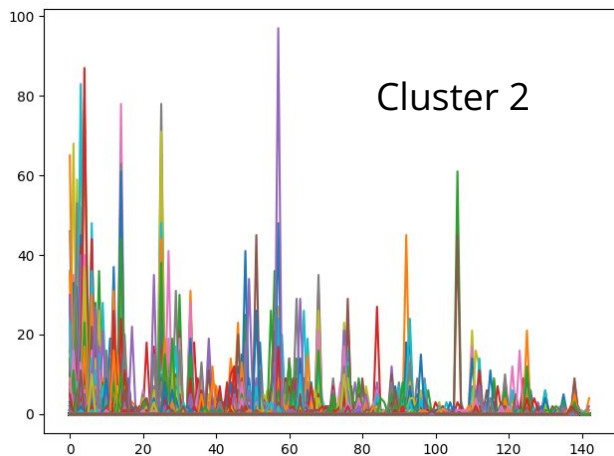
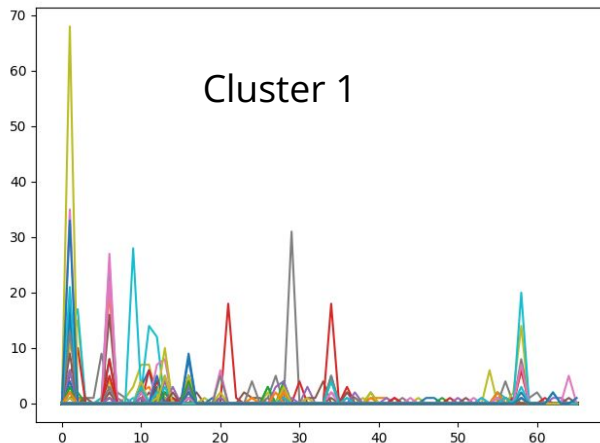
# Some Research and Methodologies with IETF Data

Priyanka Sinha, RASPRG Meeting IETF116

# Outline

- Summary from AID 2021 workshop participation and further discussions
- Relevant Methodologies for IRTF RASPRG Research
- How is IETF data valuable for Open Research?
- Summary of Hackathon

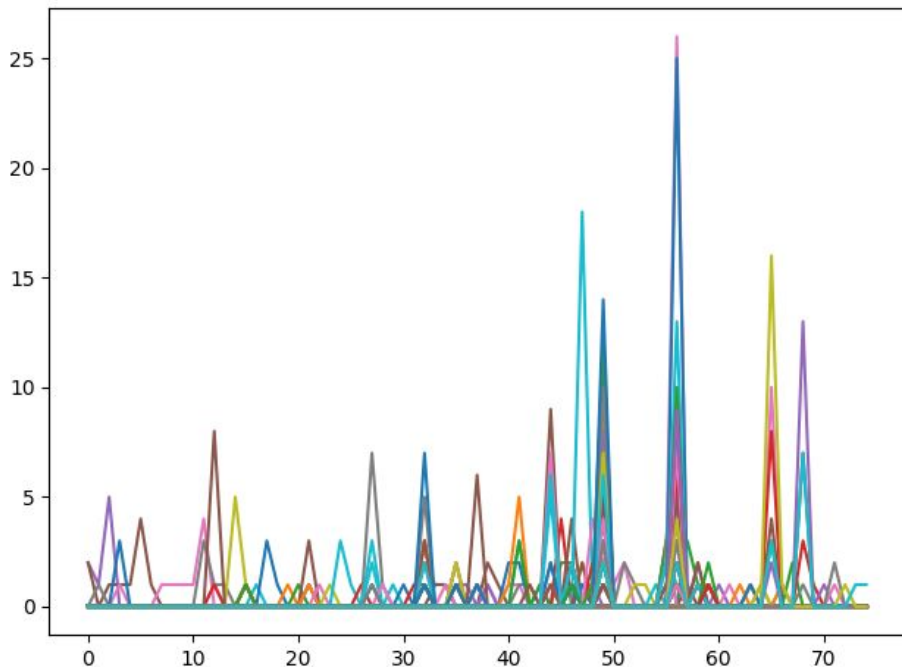
Number of emails sent by Person in the month



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

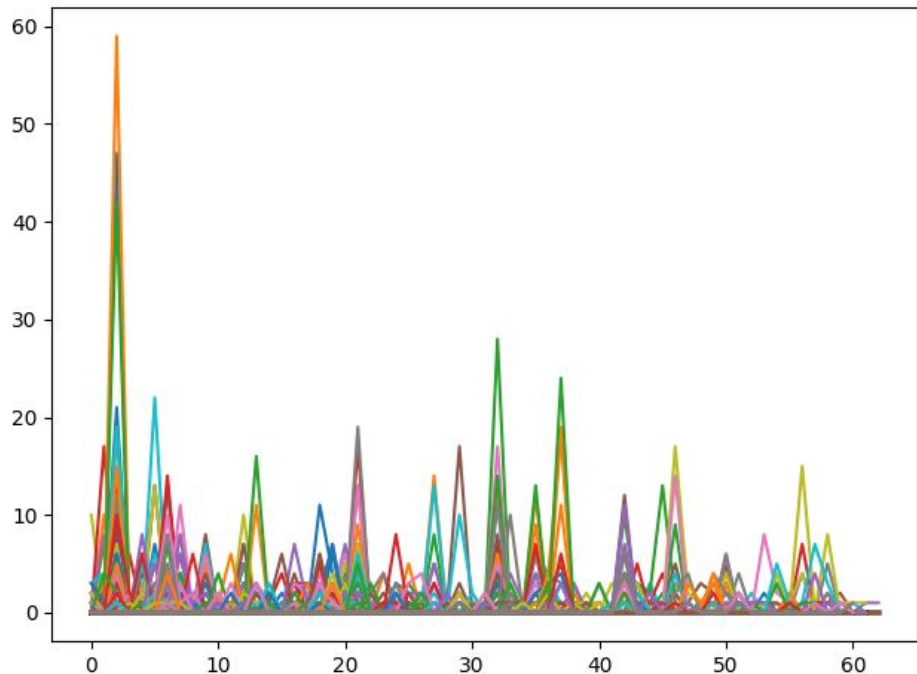
# MPLS Cluster



- Adrian Farrel [afarrel@juniper.net](mailto:afarrel@juniper.net)
- Gaurav agrawal [gaurav.agrawal@huawei.com](mailto:gaurav.agrawal@huawei.com)
- Himanshu Shah [hshah@force10networks.com](mailto:hshah@force10networks.com)
- 정태식 [cts@etri.re.kr](mailto:cts@etri.re.kr)
- 吴吉朋 [wujipeng@gmail.com](mailto:wujipeng@gmail.com)
- SP [sp543@nyu.edu](mailto:sp543@nyu.edu)
- Raghu [raghav.rao@wipro.com](mailto:raghav.rao@wipro.com)
- Arashmid Akhavain  
[arashmid@nortelnetworks.com](mailto:arashmid@nortelnetworks.com)
- [AtrJoh@netscape.net](mailto:AtrJoh@netscape.net)
- Kullberg Alan-G19424  
[alan.kullberg@motorola.com](mailto:alan.kullberg@motorola.com)
- DECRAENE Bruno RD-CORE-ISS  
[bruno.decraene@francetelecom.com](mailto:bruno.decraene@francetelecom.com)
- "Naidu, Venkata" [Venkata.Naidu@Marconi.com](mailto:Venkata.Naidu@Marconi.com)

With activity peaks aligned, these people possibly have the similar viewpoint or interest within MPLS WG. Further text mining on their content, would provide say topics and opinions for new participants at IETF to find shepherds and advocates.

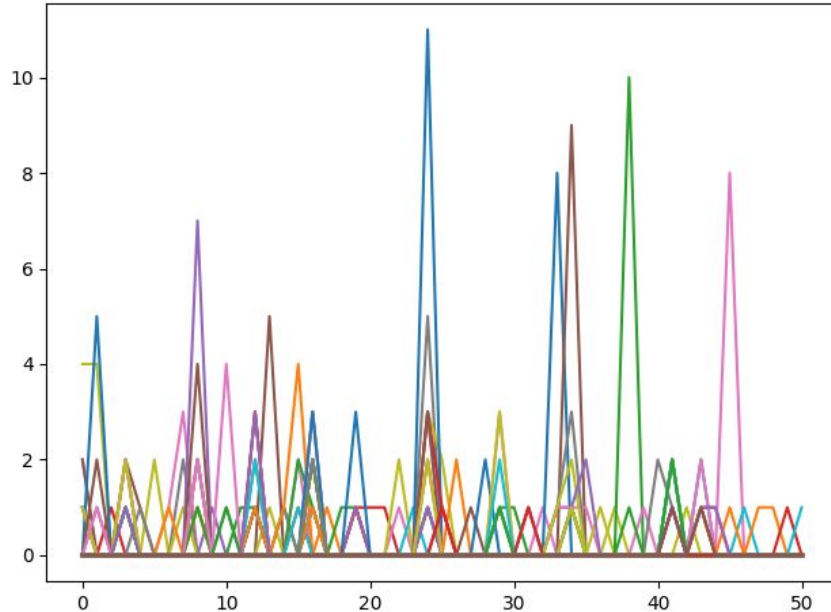
# IDR Cluster



- Tony Li [tli@juniper.net](mailto:tli@juniper.net)
- Jakob Heitz [jheitz@redback.com](mailto:jheitz@redback.com)
- Gargi Nalawade [nalawade@redback.com](mailto:nalawade@redback.com)
- Saikat Ray [ray.saikat@ericsson.com](mailto:ray.saikat@ericsson.com)
- Adrian Farrel [adrian@olddog.co.uk](mailto:adrian@olddog.co.uk)
- Dhruv Dhody [dhruv.ietf@gmail.com](mailto:dhruv.ietf@gmail.com)
- Yimin Shen [yshen@juniper.net](mailto:yshen@juniper.net)

Another example of entity disambiguation of emails and affiliations from temporal activity. I know some of these people have worked together in person earlier, even though their affiliation has changed, even last name! Likely have similar technical opinions and help IETF to speed process of consensus and maybe even adoption.

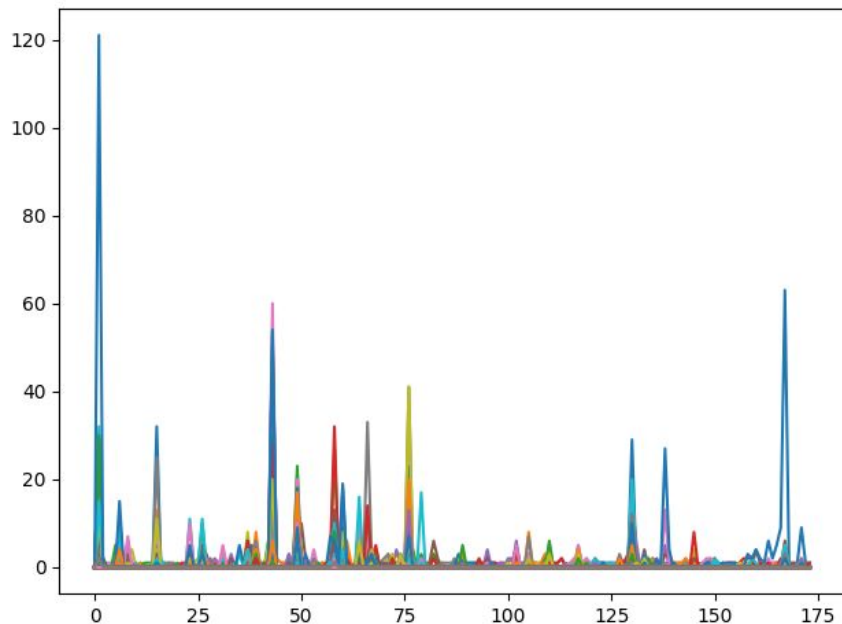
# CCAMP Cluster



- ogino [ogino@kddilabs.jp](mailto:ogino@kddilabs.jp)
- Wataru Imajuku  
[imajuku.wataru@lab.ntt.co.jp](mailto:imajuku.wataru@lab.ntt.co.jp)
- Acee Lindem [acee@cisco.com](mailto:acee@cisco.com)
- Dhruv Dhody [dhruv.ietf@gmail.com](mailto:dhruv.ietf@gmail.com)
- "Linwei (Wei)" [wei.linwei@huawei.com](mailto:wei.linwei@huawei.com)
- Don Fedyk [dwfedyk@nortel.com](mailto:dwfedyk@nortel.com)
- FEDYK Don  
[Donald.Fedyk@alcatel-lucent.com](mailto:Donald.Fedyk@alcatel-lucent.com)
- David Ward [dward@juniper.net](mailto:dward@juniper.net)
- VICTOR LOPEZ ALVAREZ  
[victor.lopezalvarez@telefonica.com](mailto:victor.lopezalvarez@telefonica.com)
- Francesco Lazzeri  
[francesco.lazzeri@ericsson.com](mailto:francesco.lazzeri@ericsson.com)

Seemingly different demographics and affiliations may have similar opinions or interests. Motivation for IETF to look beyond demographics and affiliations.

# All Combined Cluster



- Adrian Farrel - [adrian@olddog.co.uk](mailto:adrian@olddog.co.uk),  
[olddog@clara.co.uk](mailto:olddog@clara.co.uk),  
[drian-nomcom@olddog.co.uk](mailto:drian-nomcom@olddog.co.uk)
- "Dutta, Pranjal (Pranjal)"  
[pdutta@alcatel-lucent.com](mailto:pdutta@alcatel-lucent.com),  
[pranjal.dutta@alcatel-lucent.com](mailto:pranjal.dutta@alcatel-lucent.com)
- 정연쾌 [ykjeong@etri.re.kr](mailto:ykjeong@etri.re.kr)
- [E.T.Metz@telecom.tno.nl](mailto:E.T.Metz@telecom.tno.nl)
- Jakob Heitz [jheitz+041207@redback.com](mailto:jheitz+041207@redback.com)
- Balaji Pitta venkatachalapathy  
[balaji\\_pv@hotmail.com](mailto:balaji_pv@hotmail.com)

Entity disambiguation from temporal analytics helps identify the person, downstream text mining for richer analytics for IETF

# Causal Learning Dialogue Outcomes

Attribute	Predicate	Value Domain
sentiment	SENTIMENT_VERY_NEG SENTIMENT_NEG SENTIMENT_LOW_POS SENTIMENT_MEDIUM_POS SENTIMENT_HIGH_POS	< -0.2 [-0.2,0) [0,0.5) [0.5,0.8) >0.8
EMPATH	CATEGORY_ABSENT CATEGORY_LOW CATEGORY_MEDIUM CATEGORY_HIGH	=0 (0,1) [1,2) >=2

Formal method text mining to study IETF mailing list conversations, their outcomes and causes of the same. May help in speeding consensus.

12:15:46 I feel Very **elated** to wish Happy Womens Day but really **sad** when its tailored as **Happy** and **safety** Womens day (... I can understand that society is changing in a very **bad** shape. Yes the **Men** Clan is **responsible** for This addition of **Safety** Womens Day tag. I know I cant change the whole **community** overnite, but yes, I can make an attempt to remove the Tag of **Safety** and convert it INTO just **HAPPY** WOMENS DAY :)

14:28:03 **Thank** you Sujal...

14:10:07 **Nicely** written Epsy. Time for all to change. **Girls** as well. Not your dressing. Not your **fun**. Far time for us to stop depicting ourselves as delicate darlings and bring the **evil** out. **Dont** make it easy for them. Let it not be a one sided war. Enough of branding ourselves as **physically weak** & mentally **strong** humans. **mental** strength will not work here.. have your tools.. have your **guts**. Cut it out. I mean it. **Dont** wait for soldiers.. we **dont** need them.. they have their own job to do.. be you own soldier. **Happy** Womens Day..

14:27:30 **Well** said dyana. **Happy** womens day..

21:31:33 **Good** post Epsy.. I have nothing but my words to ensure that I'll be a **good** person and **respect** women all my life.. I can extend my hand to help whenever and wherever is possible... Wish you all a very **happy** Womens Day..

**IDOMESTIC\_WORK\_ABSENT &&  
MEDICAL\_EMERGENCY\_ABSENT**

**FAMILY\_ABSENT && !SENTIMENT\_POS  
##[0:3000] FAMILY\_HIGH &&  
SENTIMENT\_VERY\_POS##[0:10000]  
DOMESTIC\_WORK\_ABSENT**

+1 19:37:27 Yes I **agree** with you but I really **dont** like when it comes and lick you...!

+1 21:36:11 **ha ha ha** **thats** due to their **loving** nature. I guess :)

+6 12:12:47 Cleaning the desktop?????

+8 08:49:34 yaaa **frm** inside... :P

+64 13:12:40 what a story...**nd** people still fall for it :)

+64 14:36:22 yaa **this** the **trick**..... :P

+64 15:01:06 yeah **jus** like u fell for it :)

+64 20:27:50 yaa like **evryone** i also fell for it... :)

+65 11:27:47 u seem so **happy** that u fell for it :)

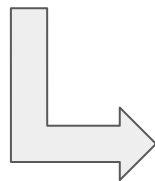
+93 19:58:05 So Hutch/Vodafone **dog** has got a new **job**... Cleaning or systems screen from inside.... :P



# Multigraph Edge Features for Graph Mining Peer Groups

- **Enron data set**

	#words	#vocabulary	EMPATH	#emails sent	#emails recvd	sentiment	#entities	#capital words
avg	1047	949	0.004	2	3	0.3	109	222
max	101549	92491	4.31	226	343	22.8	9972	21504
min	1	1	0	1	1	-4.3	0	0



	#urls	#verbs	#auxiliaries	#symbols	#numbers	#nouns	#adjectives	#adverbs	#pronouns
avg	10	141	198	18	38	337	57	32	36
max	1002	13870	19680	2189	5910	32349	5403	3420	3848
min	0	0	0	0	0	0	0	0	0

- **StackExchange data set**

	#questions	#comments	comment sentiment	popularity	EMPATH
avg	1.17	1.17	0.14	3.12	0.002
max	23	23	5.28	276	0.63
min	1	1	-1	0	0

# Findings from Computational Psycholinguistics

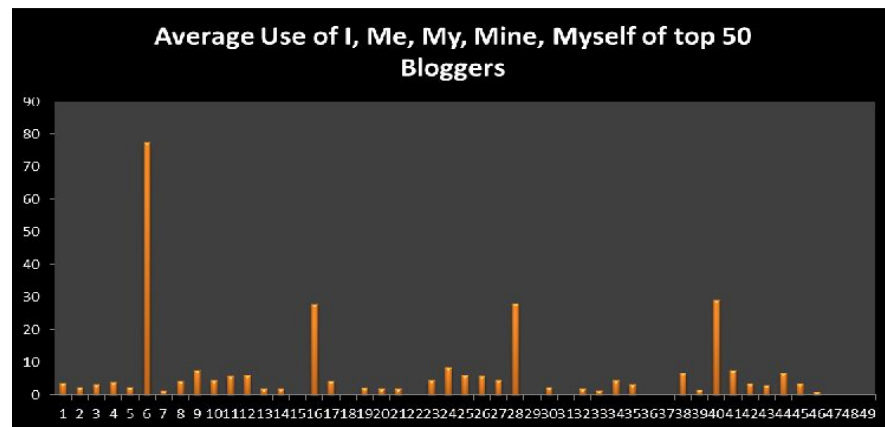
## Coverage Evaluation

Proposed Method	#Words in dictionary	#Words found in dataset	%unique words in dataset	%usage in dataset	%coverage of users
HEXACO	245	152	0.08	0.22	13
HEXACO Extension	2,108	1,999	1.07	3.95	50.18
LIWC	4,487	3,993	2.16	43.77	90.51

## Sentiment versus Traits

Corelation Between	Honesty	Emotionality	Extraversion	Agreeableness	Conscientiousness	Openness
posts versus feedback score	0.0882	0.1770	0.3362	0.3287	0.3471	0.4005
posts score versus positive emotions	-0.0914	0.2040	0.0873	0.0419	0.3476	-0.0937
Posts score versus negative emotions	-0.1316	0.1963	-0.0520	-0.0810	0.2556	0.0064
Positive emotions in posts versus feedback	0.3171	0.2700	0.3844	0.4334	0.3368	0.4689
Negative emotions in posts versus feedback	0.3147	0.5070	0.1544	0.4677	0.3845	0.3438

## LIWC Stability Observation



## Observations

- Openness is independent of use of emotive words
- Open, agreeable, extraverted, conscientious people evoke similar traits in the comments they receive from people.
- Use of 1st person personal pronouns is an indication of depression. Here the person has posted tragic short love stories.

# Why study inclusion in the IETF

- IETF is a voluntary global organization with its communication, activities recorded and available for analysis and study
- Diversity, Inclusion and Representation make the consensus process robust.
- Do consensus mechanisms depend on in-person meeting, side talks and in person advocacy?
  - This puts participants who engage in remote-only mode (perhaps due to geographical, personal constraints) at a disadvantage
- People are multi-dimensional and their sense of community may not be appropriated by their observable demographic indicators
  - Data driven identification of influence that some members may have on consensus building

# IRTF RASPRG Relevant Research Directions

- Analyzing the development of the make up of standard setting communities, their diversity, and the impact it has on standard-setting.
  - Study Contextual Integrity in terms of Group Behavior. Research to find out computational models of people's behavior from their perspective of privacy.
- Understanding the decision making processes that lead to the production of publications.
  - Identify key people who may help provide efficient and valuable consensus on drafts, proposals, rfc's and standards.

# Open Research Problems for which IETF Data is valuable

- Good data source to build a Universal Behavior Model - quantitative studies for RASP RG
  - Large engagement from various demographics.
  - Semi Formal text. Less Noisy. Not short text.
  - Many turns of dialogues. Several messages in a Thread.
  - Active open source tools - BigBang
- Provide data driven insights into personality and group behavior for a population useful in say evaluating autonomous AI for vehicles - help IETF WG with human factors insight
- Provide data driven insights into diversity, norms
- A big data resource for psychologists and cognitive scientists to test their hypotheses computationally - confirm qualitative studies with data driven insights for RASP RG

# Summary of Hackathon

# Thank You

- [https://scholar.google.co.in/citations?hl=en&user=ctlSowAAAAJ&view\\_op=list\\_works&sortby=pubdate](https://scholar.google.co.in/citations?hl=en&user=ctlSowAAAAJ&view_op=list_works&sortby=pubdate)
- <https://www.youtube.com/@PriyankaSinhaMahapatra>
- <https://www.linkedin.com/in/priyanka1982>
- [https://twitter.com/priyanka\\_iitg](https://twitter.com/priyanka_iitg)