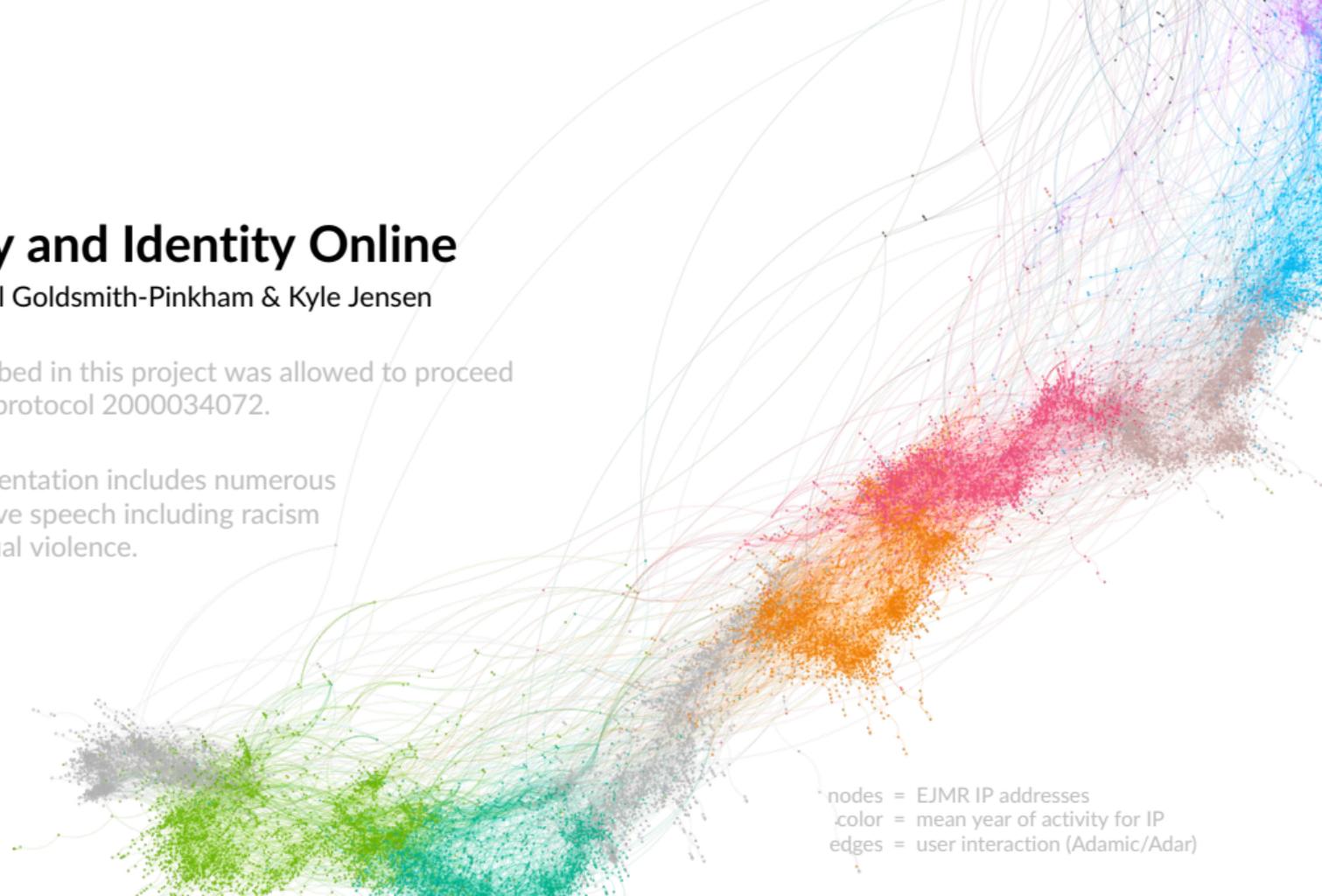


Anonymity and Identity Online

Florian Ederer, Paul Goldsmith-Pinkham & Kyle Jensen

The analysis described in this project was allowed to proceed by the Yale HRPP, protocol 2000034072.

Warning: This presentation includes numerous excerpts of offensive speech including racism and threats of sexual violence.



nodes = EJMR IP addresses
color = mean year of activity for IP
edges = user interaction (Adamic/Adar)

Description of the Project

“Economics is what economists do.”
—Jacob Viner (in Spiegel (1987), p. 814)

- Questions
 - What do economists say when they are anonymous?
 - How widespread is toxic speech in economics? Who engages in such speech?
- Econ Job Market Rumors (EJMR) is a popular anonymous message board
 - 2.5 million monthly visits (SimilarWeb, 2022)
- Statistical properties of EJMR usernames reveal IPs for the majority of posts
 - Focus today will be on **methods** of IP address identification
 - We document widespread use of EJMR even at top universities
- We use only publicly available data (Wu, 2020)

Useful Information on EJMR

ECONOMICS JOB MARKET RUMORS » ECONOMICS » ECONOMETRICS DISCUSSION



Nonlinear least squares with fixed effects

EJMR



Best paying MRM for those happy to move down

I'm an AP who's likely to get tenure pretty soon at a top 10 economics department, based on CV and signals from seniors in my dept. My goal is to move to a good-but-clearly-lower-ranked school for a big pay raise right around tenure time. I have a US preference, but am pretty locationally flexible.

Which schools should I reach out to? Do any schools have a reputation for paying a lot to recruit seniors from higher ranked schools?

EJMR



Stata graph question

This may seem like a dumb question - but is there a way of getting rid of the bluey kindergarten look of Stata graphs? How can you make graphs that look less childish? Thanks.

E 0 VOLOD 1 VLAD

!

p set scheme-
le

E 9 VOLOD 0 VLAD

our own colors for each graph (or symbols if they black and white). Or create your own color

schemes that -set scheme- should work for most

Gossip (about senior faculty moves) on EJMR

EJMR



EJMR



Zurich senior hires?

Did Zurich close any senior hires in the

8 MONTHS AGO # 04D0 QUOTE 0 VOLOD 0 VLAD

More importantly did they get rid of th

8 MONTHS AGO # 0F14 QUOTE 0 VOLOD 0 VLAD

Any News?

8 MONTHS AGO # 8D5E QUOTE 0 VOLOD 0 VLAD

bump

8 MONTHS AGO # C880 QUOTE 0 VOLOD 0 VLAD

?

8 MONTHS AGO # 60C6 QUOTE 0 VOLOD 0 VLAD

NU Macro Senior

Who did NU hire to replace MD & GL?

NU macro down to toilet?

1 YEAR AGO # 2ED9 QUOTE 0 VOLOD 0 VLAD

GL to Chicago confirmed?

1 YEAR AGO # D799 QUOTE 0 VOLOD 0 VLAD

Yes

1 YEAR AGO # 832A QUOTE 0 VOLOD 0 VLAD

GL to Chicago confirmed?

Yes

1 YEAR AGO # 2ED9 QUOTE 0 VOLOD 0 VLAD

offers

is year. Does anyone know who

LAD

!

LAD

!

LAD

!

LAD

!

LAD

!

Racism on EJMR

If your white and you ain't got least three kids your a traiter to your country

Economist
de54 thats how come them mexicans is sque

Economist ea19 There are projections saying that in 10 9 YEARS AGO # QUOTE 1 VOLOD 2 VLAD

Economist
ea19 Then they'll take back Texas using the
9 YEARS AGO # QUOTE 6 VOL0 0 VLAD

Economist
de54 not if white folks gets back to fukin an
9 YEARS AGO # QUOTE 2 VOLOD 3 VLAD

Economist
ea19 ^ Good luck. Every mexican I've met has
laws are gearing towards allowing a str
have 2 to 6 kids to keep up with the one

So tired of seeing N_____S
in my neighborhood!

18 HOURS AGO # DB0D QUOTE 1 VOLOD 0 VLAD

Just run them over and make it look like an accident. Nobody cares about a dead N R.

17 HOURS AGO # E2DE QUOTE 1 VOLOD 0 VLAD



CITI CAREER
OPPORTUNITIES
jobs.citi.com

The biggest enemies of America are:

italOne

indecareers.com

Economics-specific Racism on EJMR

EJMR



Lawrence Katz's students list

That is a long list...

https://scholar.harvard.edu/files/lkatz/files/lk_students_table_dec_2022.pdf

2 MONTHS AGO # E7B2 QUOTE 1 VOLOD 1 VLAD

Ok, but what

2 MONTHS AGO

obviously, st
to a qje. why

2 MONTHS AGO

obvi
fasti

wrong the fastest route to a qje is to grift and be black

2 MONTHS AGO # 4914 QUOTE 1 VOLOD 0 VLAD

ECONOMICS JOB MARKET RUMORS » JOB MARKET » FINANCE JOB RUMORS

YOUR TRIP TO WASHINGTON,
D.C. IS WAITING.

Pick up where you left off.

BOOK NOW

UNITED

Most toxic department in finance field

Economist

f14c

The toxicity level of a department is proportional to the number of Cheanks in the department

9 MONTHS AGO # QUOTE 1 VOLOD 2 VLAD

obviously, students at harvard know what is the easiest and fastest route to a qje. why is this surprising?

wrong the fastest route to a qje is to grift and be black

2 MONTHS AGO # 4914 QUOTE 1 VOLOD 0 VLAD

"Why do you hate intelligent successful women? What is your fear?"

9 MONTHS AGO # QUOTE 0 VOLOD 0 VLAD

Sexism and Misogyny on EJMR

Nothing can prepare you for how awful working with women is

Astonishing. Have you told Mother?

3 MONTHS AGO # 40B0 QUOTE 3 VOLOD 4 VLAD

Things were WAY better when women focused on rearing children and feeding their husbands.

3 MONTHS AGO # 453C QUOTE 8 VOLOD 6 VLAD

And slaughtering kids

Things were WAY better when women focused on rearing children and feeding their husbands.

3 MONTHS AGO # 0A18 QUOTE 1 VOLOD 1 VLAD

If you run into someone in the morning, you ran into someone who is sexist. You run into people who are emotional, selfish and incompetent all day, you're the emotional, selfish and incompetent one.

3 MONTHS AGO # 3A62 QUOTE 3 VOLOD 10 VLAD

One of our female APs parties all the time. I mean, she goes to the bar pretty often. Not to embarrass her, I stopped going to a bar where she patrons often (she didn't see me). She is very revealing too. Not sure about other female APs or male APs. I don't think she will publish up to standard. But she will almost certainly get tenure if the university doesn't want a lawsuit.

4 YEARS AGO # 07BA QUOTE 0 VOLOD 2 VLAD

My dad doesn't do any research.

Dad ++

Hire whoever has the biggest booooobs.

all girls? no diversity?

Columbia flyouts: [REDACTED], [REDACTED], [REDACTED],

[REDACTED], [REDACTED]

9 YEARS AGO # AE17 QUOTE 4 VOLOD 10 VLAD

I should work 18 hours a week to provide yourself a nice office to save your money on luxury.

People actually think like that. They have different norms about how much work is "enough". They should be avoided like the plague imo.

It is a competition, you realize that, right?

Racist and Sexist Discussion of Job Market Candidates

NYU Job Market Candidates 2022-2023

<https://as.nyu.edu/departments/econ/job-market/candidates.html>

Good luck to them!

6 MONTHS AGO # DB39 QUOTE VOLOD 0 VLAD

[REDACTED] got the biggest nose

6 MONTHS AGO # D7F1 QUOTE VOLOD 0 VLAD

Only 1 Asian out of 12?!

6 MONTHS AGO # 7A0F QUOTE VOLOD 0 VLAD

Weak

6 MONTHS AGO # 3B52 QUOTE VOLOD 0 VLAD

EJMR



Are [REDACTED] and [REDACTED] in a secret same-sekhs love-hayte relationship?

Let's move our eyes from [REDACTED]
[REDACTED] to USC Zhanettes on the job market.

They look so good omg. I am coming. I am licking the screen

4 MONTHS AGO # 3FA8 QUOTE 1 VOLOD 0 VLAD

Another week cohort from NYU....

Who writes such things?

Economist
e9dc

people from Hong Kong are the worst cheaters and rent seekers.

I lower the grades of all Mainland Chinese students. Not TW or HKers
I look at the names and they way they are romanized.
Yale System Cantonese name e.g. Wong Tai Gok = Hong Konger -> B+ becomes A-
Wade Giles e.g. Wang Teng-Hui = Taiwanese -> B+ becomes A-
Pinyin e.g. Wang Xiaodiao = Mainlander -> B+ becomes C+
I think that's only fair and compensates for the likelihood that they cheated
their way in. Also, I just don't like them.

You're working too hard. If their English is good, they are from Hong Kong. If their English is just okay, they are from Taiwan. If the only reason they passed the TOEFL is because they had another person take their exam and they speak no English at all, they are a mainlander.

3 YEARS AGO # QUOTE 0 VOLOD 0 VLAD

Economist
a483

If it is black guys on white girl it is a cultural diversity training, proper to show on TV and in cinema.

If it is white guys on black girl it is a hate crime, white guys go to prison, black girl get multimillion legal settlement.

8 YEARS AGO # QUOTE 3 VOLOD 1 VLAD

Heil Merkler!

11 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
8fed

Its about ching chong taking bubba's job and bubba putting on a white pointy hood in response.

7 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
da44

Rapefugees Welcome!!!!

- Merkel

6 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
a483

trenchant insight, bro

I just notices there is practically no black nudity on TV anymore, unless it is BBC, while there are plenty of white pussy everywhere wide open.

8 YEARS AGO # QUOTE 0 VOLOD 1 VLAD

Economist
577b

I hate hot weather. South Korea looks nice - no hotter than 23 degrees C. Chile is good too. I don't want to work in some miserable hot, sticky jungle or desert place.

Chile gets warm. And bitches are fugly.

If you don't like hot weather you should consider Moscow, Russia.

8 YEARS AGO # QUOTE 1 VOLOD 2 VLAD

Economist
6b12

after all the people giving him blow jobs on EJMRs you would have expected great things.

9 YEARS AGO # QUOTE 2 VOLOD 3 VLAD

Who writes such things?

Economist
e9dc

people from Hong Kong are the worst cheaters and rent seekers.

I lower the grades of all Mainland Chinese students. Not TW or HKers
I look at the names and they way they are romanized.
Yale System Cantonese name e.g. Wong Tai Gok = Hong Konger → B+ becomes A-
Wade Giles e.g. Wang Teng-Hui = Taiwanese → B+ becomes A-
Pinyin e.g. Wang Xiaodiao = Mainlander → B+ becomes C+
I think that's only fair and compensates for the likelihood that they cheated
their way in. Also, I just don't like them.

You're working too hard. If their English is good, they are from Hong Kong. If their English is just okay, they are from Taiwan. If the only reason they passed the TOEFL is because they had another person take their exam and they speak no English at all, they are a mainlander.

3 YEARS AGO # QUOTE 0 VOLOD 0 VLAD

Economist
a483

If it is black guys on white girl it is a cultural diversity training, proper to show on TV and in cinema.

If it is white guys on black girl it is a hate crime, white guys go to prison, black girl get multimillion legal settlement.

8 YEARS AGO # QUOTE 3 VOLOD 1 VLAD

Economist
0fee

Heil Merkler!

11 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
8fed

Its about ching chong taking bubba's job and bubba putting on a white pointy hood in response.

7 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
da44

Rapefugees Welcome!!!!

- Merkel

6 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
a483

trenchant insight, bro

I just notices there is practically no black nudity on TV anymore, unless it is BBC, while there are plenty of white pussy everywhere wide open.

8 YEARS AGO # QUOTE 0 VOLOD 1 VLAD

Economist
577b

I hate hot weather. South Korea looks nice - no hotter than 23 degrees C. Chile is good too. I don't want to work in some miserable hot, sticky jungle or desert place.

Chile gets warm. And blettes are fugly.

If you don't like hot weather you should consider Moscow, Russia.

8 YEARS AGO # QUOTE 1 VOLOD 2 VLAD

Economist
6b12

after all the people giving him blow jobs on EJMRs you would have expected great things.

9 YEARS AGO # QUOTE 2 VOLOD 3 VLAD

IP addresses at Harvard, Stanford, Yale, Chicago ...

Who writes such things?

Economist
e9dc

people from Hong Kong are the worst cheaters and rent seekers.

I lower the grades of all Mainland Chinese students. Not TW or HKers
I look at the names and they way they are romanized.
Yale System Cantonese name e.g. Wong Tai Gok = Hong Konger → B+ becomes A-
Wade Giles e.g. Wang Teng-Hui = Taiwanese → B+ becomes A-
Pinyin e.g. Wang Xiaodiao = Mainlander → B+ becomes C+
I think that's only fair and compensates for the likelihood that they cheated
their way in. Also, I just don't like them.

You're working too hard. If their English is good, they are from Hong Kong. If their English is just okay, they are from Taiwan. If the only reason they passed the TOEFL is because they had another person take their exam and they speak no English at all, they are a mainlander.

3 YEARS AGO # QUOTE 0 VOLOD 0 VLAD

Economist
a483

If it is black guys on white girl it is a cultural diversity training, proper to show on TV and in cinema.

If it is white guys on black girl it is a hate crime, white guys go to prison, black girl get multimillion legal settlement.

8 YEARS AGO # QUOTE 3 VOLOD 1 VLAD

Economist
0fee

Heil Merkler!

11 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
8fed

Its about ching chong taking bubba's job and bubba putting on a white pointy hood in response.

7 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
da44

Rapefugees Welcome!!!!

- Merkel

6 YEARS AGO # QUOTE 1 VOLOD 0 VLAD

Economist
a483

trenchant insight, bro

I just notices there is practically no black nudity on TV anymore, unless it is BBC, while there are plenty of white pussy everywhere wide open.

8 YEARS AGO # QUOTE 0 VOLOD 1 VLAD

Economist
577b

I hate hot weather. South Korea looks nice - no hotter than 23 degrees C. Chile is good too. I don't want to work in some miserable hot, sticky jungle or desert place.

Chile gets warm. And bitches are fugly.

If you don't like hot weather you should consider Moscow, Russia.

8 YEARS AGO # QUOTE 1 VOLOD 2 VLAD

Economist
6b12

after all the people giving him blow jobs on EJMRs you would have expected great things.

9 YEARS AGO # QUOTE 2 VOLOD 3 VLAD

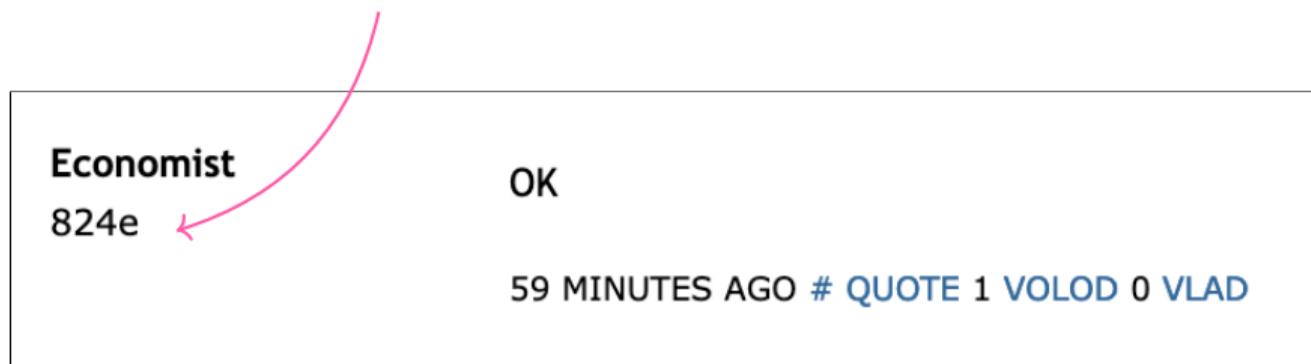
IP addresses at Harvard, Stanford, Yale, Chicago ... and the NBER HQ at 1050 Mass Ave

Username Allocation on EJMR

- Each topic is assigned a **topic id**
 - <https://www.econjobrumors.com/topic/right-vs-left-wing-dictatorships>
 - <https://www.econjobrumors.com/topic/1127272>
-
- The diagram illustrates the structure of a URL for a topic on the EJMR website. The URL is shown as `https://www.econjobrumors.com/topic/1127272`. A blue curved arrow points from the text "topic id" to the number `1127272`, indicating it is a numerical identifier. Another blue curved arrow points from the text "slug" to the word `right-vs-left-wing-dictatorships`, indicating it is a descriptive string used in the URL.
- An incrementing counter common to WordPress sites

Username Allocation on EJMR

- Each topic is assigned a **topic id**
 - <https://www.econjobrumors.com/topic/right-vs-left-wing-dictatorships>
 - <https://www.econjobrumors.com/topic/1127272>
- Each post is assigned a **username**, which is topic-specific



Username Allocation on EJMR

- Each topic is assigned a **topic id**
 - <https://www.econjobrumors.com/topic/right-vs-left-wing-dictatorships>
 - <https://www.econjobrumors.com/topic/1127272>
- Each post is assigned a **username**, which is topic-specific
- **Username** was characters 10-13 of the SHA-1 hash of **topic ID** and **IPv4 address**
 - Hash **did not use a salt** (a random secret) and was in plain sight for over a decade
- E.g. **824e** – a 4-digit hexadecimal

Topic ID: **1127272**

IPv4 Address: **130.132.173.94**

SHA-1 Input: **1127272130.132.173.94**

SHA-1 Output: **eae3d51bc824e273e203b9fbfb608828431a6d48**

Username: **824e**

From usernames to IP addresses

- SHA-1 hash is one-way. It cannot be reversed.
- But we can try every possible IP address to see if some of them match:

11272720.0.0.0	→	59d5e2cc45a94aab27a1c5e69644783d9f16726a
11272720.0.0.1	→	d77bff2cf9a90e8e2053b9b5d28847a1cf919510
⋮	⋮	⋮
1127272130.132.173.94	→	eae3d51bc824e273e203b9fbfb608828431a6d48
⋮	⋮	⋮
1127272255.255.255.255	→	0ef63f5c3d79e2fce6a2270dc52f88118a2a3949

- Record all IP addresses where observed topic-username matches the hash
 - Feasible on GPUs
 - $695,364 \text{ topics} \times 2^{32} \text{ possible IPv4 addresses} \approx 3 \text{ quadrillion hashes}$
 - Returns a set of matching IPs per topic-username: 65,536 in expectation

First post:

IP
topic ID
username

130.132.173.94
1127272
824e

My IP

130.132.173.94

Second post:

IP
topic ID
username

130.132.173.94
1127329
607e

Economist
824e

OK

59 MINUTES AGO # QUOTE 1 VOLOD 0 VLAD

Economist
607e

OK

59 MINUTES AGO # QUOTE 1 VOLOD 0 VLAD

hash
inversion

Matching IPs for 1127272/824e

65,668
IPs

0.1.96.120	0.2.213.60
0.3.5.107	0.4.41.216
0.4.159.248	0.5.109.255
0.6.29.0	0.7.104.143
.	.
255.248.137.217	255.249.166.138
255.250.174.16	255.252.53.219
255.253.207.68	255.254.36.217
255.254.220.185	255.254.220.254

0.0.229.36	0.1.4.220
0.2.255.120	0.3.115.129
0.4.27.95	0.5.147.178
0.7.209.187	0.8.254.155
.	.
255.246.223.98	255.247.105.179
255.247.157.134	255.250.145.18
255.252.29.118	255.254.21.159
255.255.56.111	255.255.101.89

65,907
IPs

Which IPs occur in both sets?

First post:

My IP

Second post:

IP
topic ID
username

130.132.173.94
1127272
824e

130.132.173.94

IP
topic ID
username

130.132.173.94
1127329
607e

Economist
824e

OK

59 MINUTES AGO # QUOTE 1 VOLOD 0 VLAD

Economist
607e

OK

59 MINUTES AGO # QUOTE 1 VOLOD 0 VLAD

hash
inversion

Matching IPs for 1127272/824e

65,668
IPs

0.1.96.120	0.2.213.60
0.3.5.107	0.4.41.216
0.4.159.248	0.5.109.255
0.6.29.0	0.7.104.143
.	.
255.248.137.217	255.249.166.138
255.250.174.16	255.252.53.219
255.253.207.68	255.254.36.217
255.254.220.185	255.254.220.254

0.0.229.36	0.1.4.220
0.2.255.120	0.3.115.129
0.4.27.95	0.5.147.178
0.7.209.187	0.8.254.155
.	.
255.246.223.98	255.247.105.179
255.247.157.134	255.250.145.18
255.252.29.118	255.254.21.159
255.255.56.111	255.255.101.89

65,907
IPs

Which IPs occur in both sets?
Just one: 130.132.173.94

Probabilistic Identification of Active IPs

- Identifying true IP is possible because
 1. True IP is *always* present in matching set
 2. Other “noise” IPs are i.i.d. **uniformly** over IP space due to SHA-1 avalanche property

True IPs will show up much more often than noise IPs!

- Statistical properties
 - Probability of a “noise” IP appearing in a topic follows a **hypergeometric distribution**
 - Number of times a “noise” IP appears in a week follows a **Poisson binomial distribution**

An Attribution Example

What is the IP address
for this post with
topic id = 175901 and
username = 6b42?

The screenshot shows a forum post on the 'Economics Job Market Rumors' board. The post has the following details:

- Topic:** ECONOMICS JOB MARKET RUMORS » OFF TOPIC » TRASH
- Post Content:** One of the worse things about industry is having to work with women
- User Information:** Economist, 6b42
- Post Text:** I cannot stress this enough. You know the kind of women that are terrible at quant work and just will full you over to advance their career. Even worse if she's a mother as she has even more incentive to move up the ladder without having any noticeable skill or quality.
YOU HAVE BEEN WARNED
- Timestamp:** 6 YEARS AGO
- Interaction:** # QUOTE 33 JAB 11 NO JAB
- Follow-up Post:** Why didn't you hit it and quit when you had the chance?
6 YEARS AGO # QUOTE 0 JAB 0 NO JAB
- Final Post:** where the whaaaaaaaaaaaaaaaaambulance when op needs it?!!
6 YEARS AGO # QUOTE 6 JAB 12 NO JAB

A red arrow points from the question in the sidebar to the user information 'Economist 6b42' in the first post.

An Attribution Example

- We start with 2^{32} possible IP addresses.
- The hash inversion narrows it down to 65,385 matching IP addresses for this post.
- How many of these 65,385 IPs explain **other** topic-usernames in a 7-day window?

# other explained (t, u)	p-value	# IP addresses
0	1.00000000	58,367

An Attribution Example

- We start with 2^{32} possible IP addresses.
- The hash inversion narrows it down to 65,385 matching IP addresses for this post.
- How many of these 65,385 IPs explain **other** topic-usernames in a 7-day window?

# other explained (t, u)	p-value	# IP addresses
0	1.00000000	58,367
1	0.10694695	6,627

about 6,605
expected by chance

An Attribution Example

- We start with 2^{32} possible IP addresses.
- The hash inversion narrows it down to 65,385 matching IP addresses for this post.
- How many of these 65,385 IPs explain **other** topic-usernames in a 7-day window?

# other explained (t, u)	p-value	# IP addresses
0	1.00000000	58,367
1	0.10694695	6,627
2	0.00592438	378



about 373
expected by chance

An Attribution Example

- We start with 2^{32} possible IP addresses.
- The hash inversion narrows it down to 65,385 matching IP addresses for this post.
- How many of these 65,385 IPs explain **other** topic-usernames in a 7-day window?

# other explained (t, u)	p-value	# IP addresses
0	1.00000000	58,367
1	0.10694695	6,627
2	0.00592438	378
3	0.00022049	12

about 12
expected by chance

An Attribution Example

- We start with 2^{32} possible IP addresses.
- The hash inversion narrows it down to 65,385 matching IP addresses for this post.
- How many of these 65,385 IPs explain **other** topic-usernames in a 7-day window?

# other explained (t, u)	p-value	# IP addresses
0	1.00000000	58,367
1	0.10694695	6,627
2	0.00592438	378
3	0.00022049	12
4	6.167×10^{-6}	0

about 0.39
expected by chance

An Attribution Example

- We start with 2^{32} possible IP addresses.
- The hash inversion narrows it down to 65,385 matching IP addresses for this post.
- How many of these 65,385 IPs explain **other** topic-usernames in a 7-day window?

# other explained (t, u)	p-value	# IP addresses
0	1.00000000	58,367
1	0.10694695	6,627
2	0.00592438	378
3	0.00022049	12
4	6.167×10^{-6}	0
5	1.380×10^{-7}	0



about 0.0089
expected by chance

An Attribution Example

- We start with 2^{32} possible IP addresses.
- The hash inversion narrows it down to 65,385 matching IP addresses for this post.
- How many of these 65,385 IPs explain **other** topic-usernames in a 7-day window?

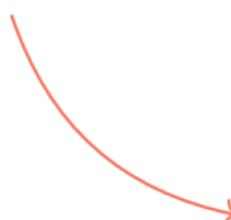
# other explained (t, u)	p-value	# IP addresses
0	1.00000000	58,367
1	0.10694695	6,627
2	0.00592438	378
3	0.00022049	12
4	6.167×10^{-6}	0
5	1.380×10^{-7}	0
⋮	⋮	⋮

An Attribution Example

- We start with 2^{32} possible IP addresses.
- The hash inversion narrows it down to 65,385 matching IP addresses for this post.
- How many of these 65,385 IPs explain **other** topic-usernames in a 7-day window?

# other explained (t, u)	p-value	# IP addresses
0	1.00000000	58,367
1	0.10694695	6,627
2	0.00592438	378
3	0.00022049	12
4	6.167×10^{-6}	0
5	1.380×10^{-7}	0
⋮	⋮	⋮
102	2.253×10^{-258}	0
103	2.072×10^{-267}	1

Very much not
expected by chance



IP Attribution as an Optimization Problem

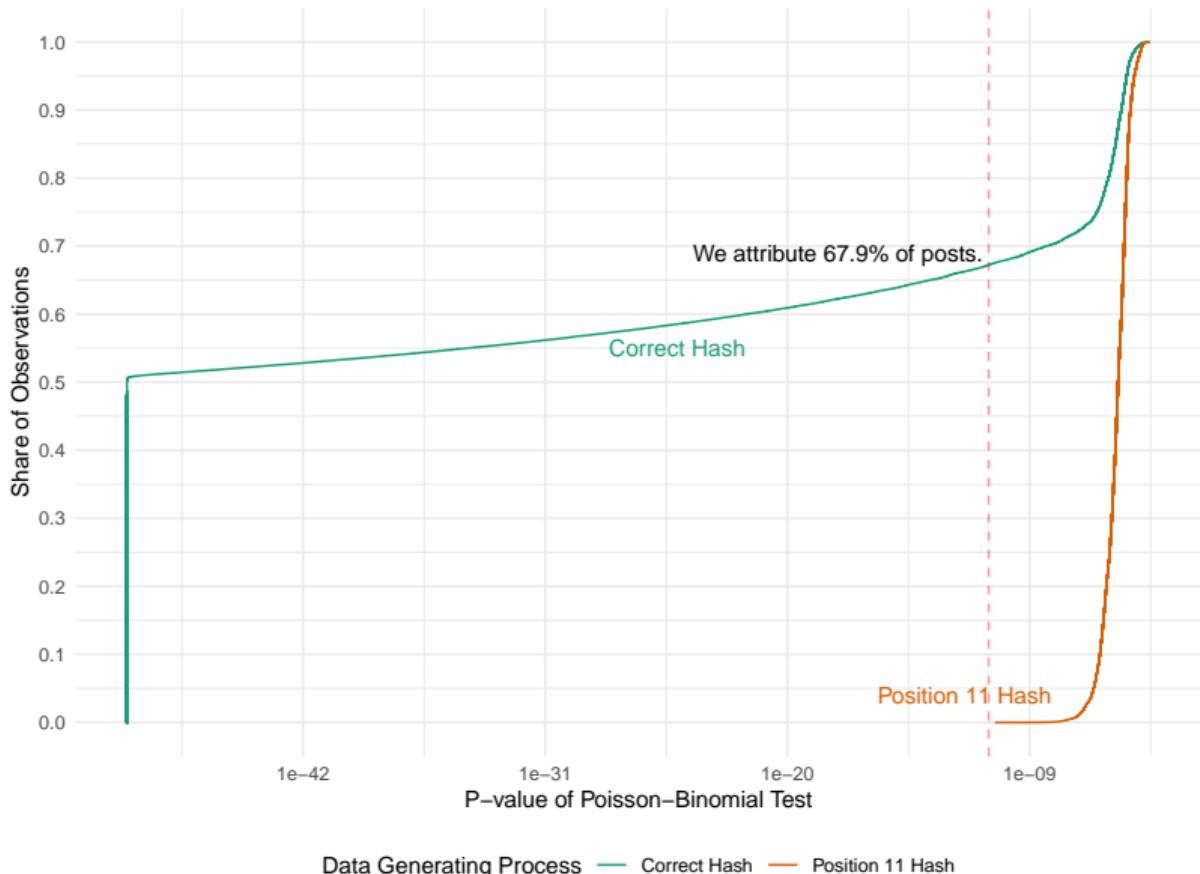
- Find a small set of these active IPs that explains the observed data.
 - Subject to locality and significance constraints
 - *Definitely has errors.* These are estimable and in the paper.
- Simple intuition of this optimization approach
 - 65k IP addresses can explain any given post.
 - One IP explains many other posts that week.
 - What is the likely origin IP of this post?
 - It's probably this highly explanatory IP.



This is essentially an argument that
number of users \ll 4.3B possible IP addresses

Attribution with Correct and Incorrect Hashes

- Using any incorrect substring of the hash only generates “noise” IPs
- Determine p-value threshold p^* by comparing distributions of minimum p-values
- Only assign IPs to posts with minimum p-value $< p^* \approx 10^{-11}$



Robustness Checks for Attribution

- Incorrect hashing set ✓
 - Using the (incorrect) position-11 hashing set, **none** of the roughly 7.1 million posts observed on EJMR are attributed an IP address.
- Bogon addresses ✓
 - There are nearly 600 million bogon addresses which occupy 13.8% of the entire IPv4 address space. These cannot post to EJMR but could be attributed to posts by mistake.
 - Our process attributes **zero** posts to bogon IP addresses.
- Time pattern of posters ✓
 - IP addresses post during the **standard work and day time hours** of their geolocation.
- Language of posters ✓
 - The dominant non-English language of the country of origin of the IP address is the **country's native language**.

From Methods to Results

What tools does our analysis give us?

1. Panel dataset linking poster IPs across topics and time
2. Geolocation of posters (high quality at state and country level, can go down to city)
3. Information on ISPs (universities, organizations, corporations, hotels)

What questions are we asking?

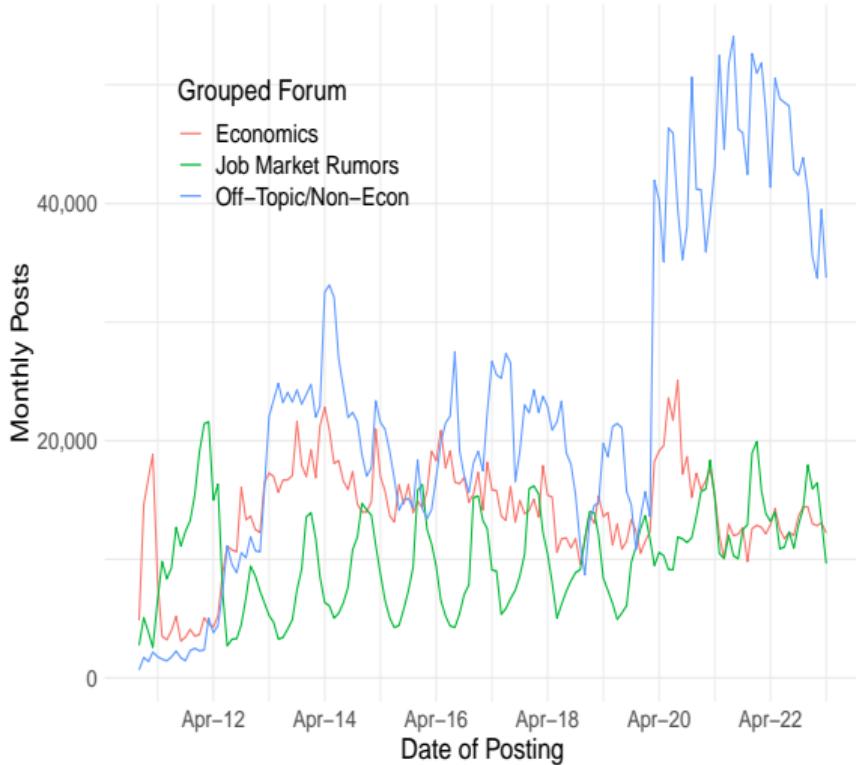
- Who are the posters on this site?
 - ~~“It’s just people at lower ranked universities.”~~
 - ~~“It’s just grad students.”~~
- Is the toxicity of the site widespread?
 - ~~“It’s just a few bad apples.”~~
- Are there two dialogues (toxic and professional)?
Are they separate? Within people? Within topic?
- Is there actually valuable inside information?
- Does EJMR make people more toxic?
- Are hundreds of thousands of visitors just paying attention to what a few people post?

Attribution and Geolocation

- 7.1 million posts in total → 4.7 million posts (66.1%) with attributed IP address
- 47,630 distinct IP addresses attributed to posts
 - These are the most frequent posters, but ...
 - ... there are many more infrequent posters, and ...
 - ... there are even more viewers (ratio of roughly 70 views to 1 post).
- Vast majority from countries with top research institutions in economics & finance
 - US (61.9%)
 - Canada (8.3%), United Kingdom (5.5%)
 - Australia (2.4%), Germany (2.2%), Hong Kong (1.9%), Italy (1.6%), France (1.5%)
 - Remaining share of geolocated posts (13.6%) from rest of the world

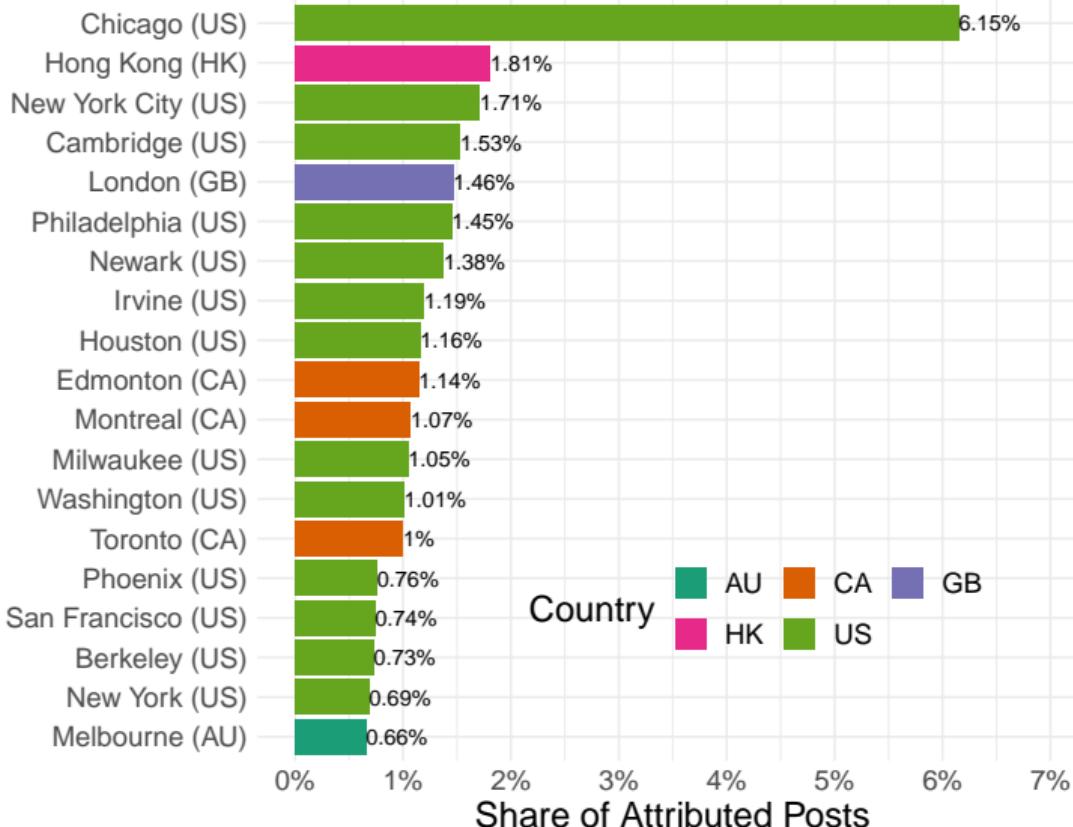
Time Pattern of Posts

- 7.1 million posts since December 2010
 - Average 70,000 monthly posts & 1,100 monthly unique IPs as of 2022
- Steady increase over time, but large traffic increase during COVID-19
 - Primarily driven by tripling in U.S.
 - Other countries experience more temporary increases
 - Very large rise in off-topic forum posts
- Strong cyclicity of job market posts
 - Disruption of job market cyclicity starting in 2021



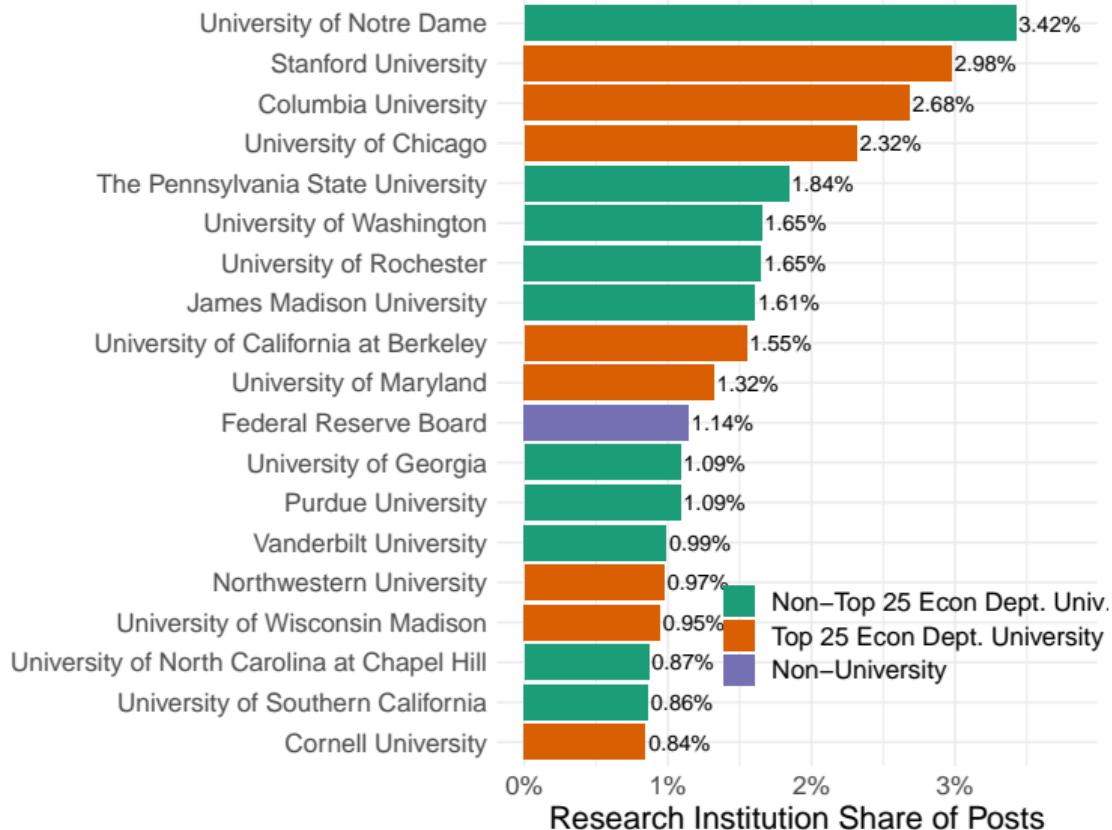
Posts by City

- Majority of posts come from large cities in the US
 - Chicago, New York, Philadelphia
- Some fraction from large cities outside the US
 - Hong Kong, London, Montreal, Toronto
- Smaller US cities with leading research institutions
 - Cambridge, Berkeley



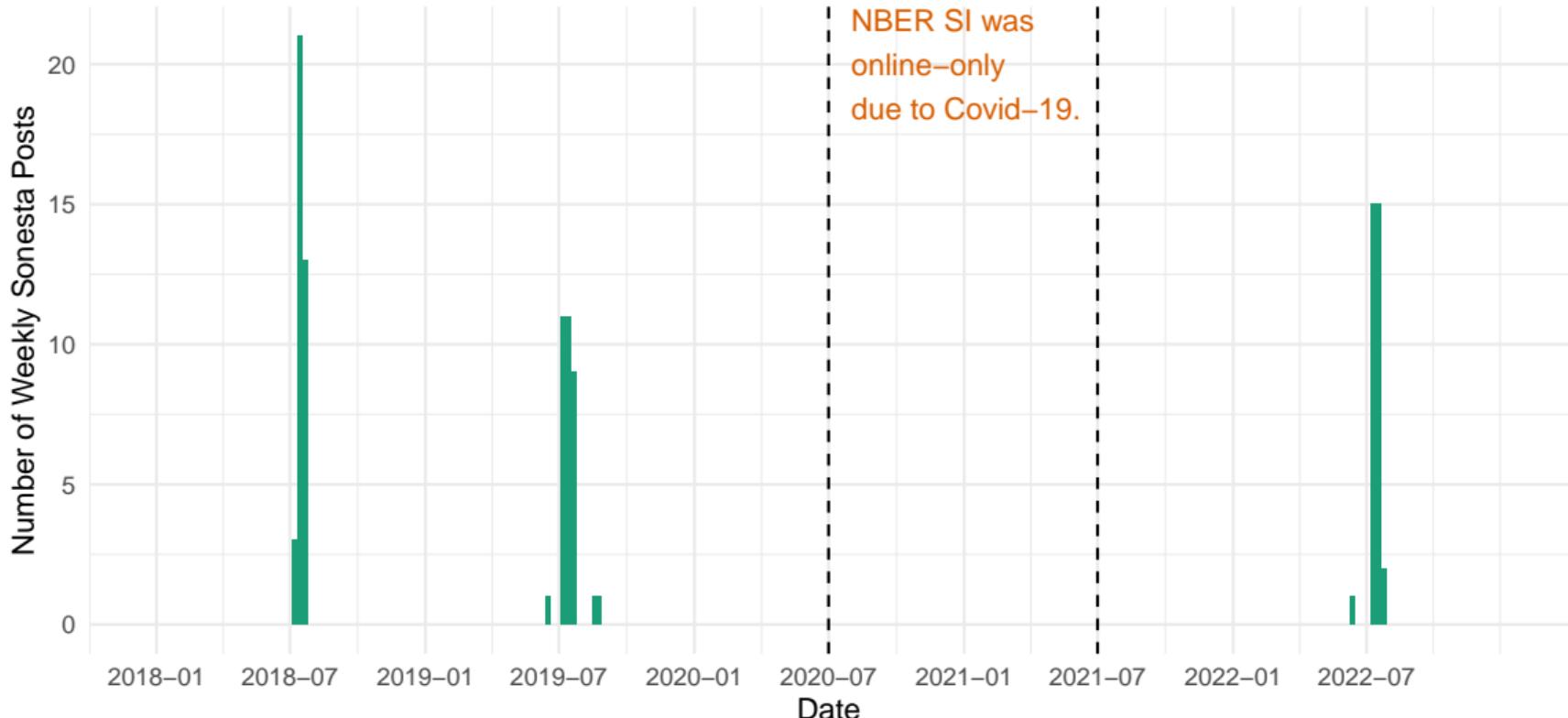
University IP Addresses on EJMR

- 10.9% of allocated posts originate from IP addresses of universities or research institutions
- Contributors are using university networks to post on EJMR
- Posts come from top US universities (top 25 econ departments)



Are these the elites of economics?

Posts concentrate in July at the Royal Sonesta



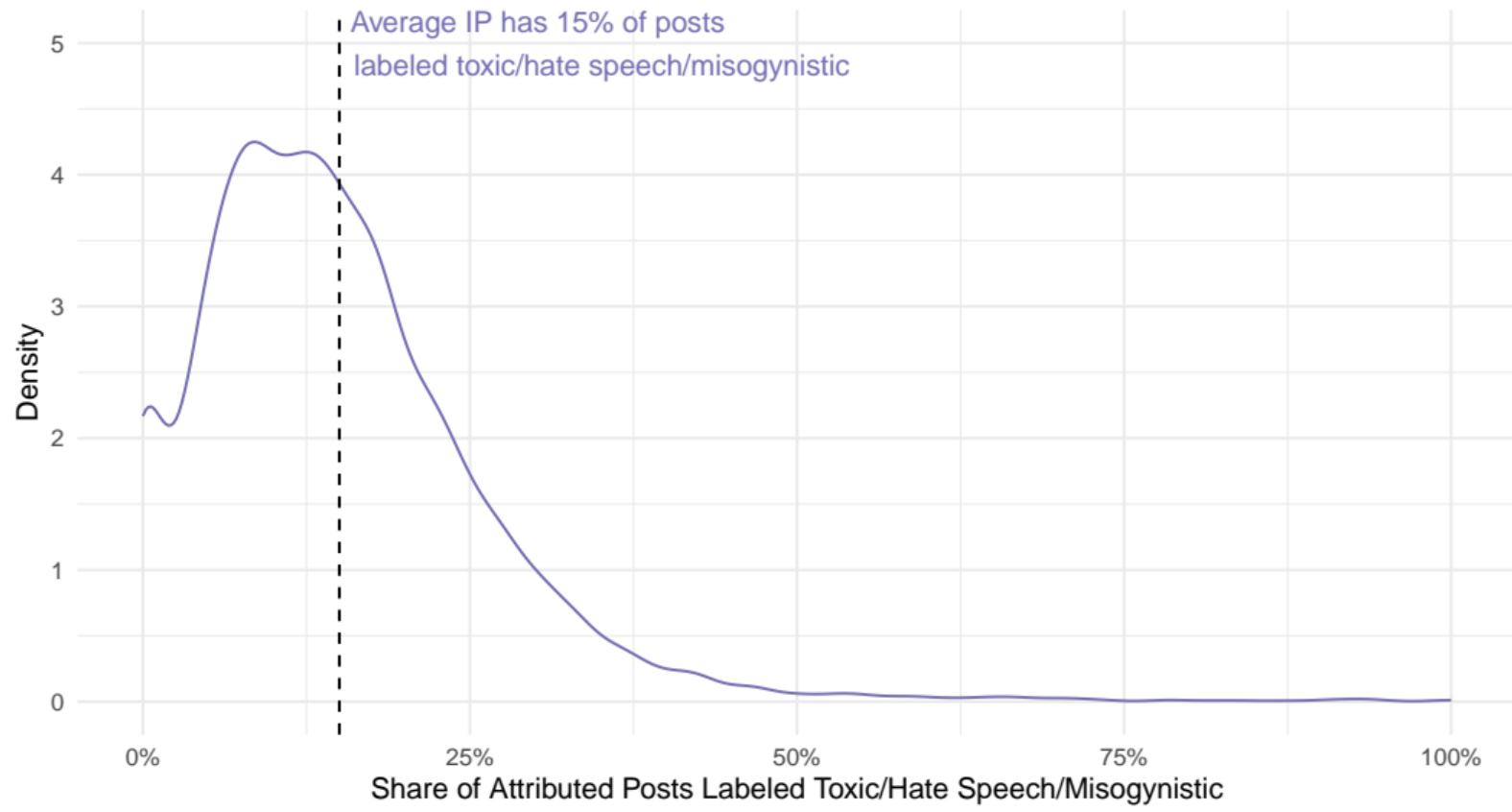
NBER SI was
online-only
due to Covid-19.

Which universities do university IPs talk about?

University ISP	Share of posts that mention university									
	Harvard	MIT	Stanford	Berkeley	UChicago	Yale	NYU	NWU	Columbia	UPenn
Harvard	7.9	9	5.2	1.4	3.7	2	2.1	0.9	1	1.2
MIT	4.7	9.8	6	0.9	2.6	0.4	2.6	1.7	2.1	1.3
Stanford	4.4	6.4	7.4	1.7	4.5	1.2	1.9	1.4	1.3	1.3
UC Berkeley	1.6	3.7	1.7	4.2	1.8	1.2	1.8	0.8	0.6	0.9
UChicago	2.1	4.8	1.4	0.7	8.3	0.8	1.6	0.7	0.5	1
Yale	1.5	3.8	0.9	0.7	1.8	3.4	1.3	0.5	0.4	1
NYU	2.5	4.6	3.1	0.7	1.9	1.1	5.8	1	1.1	2.4
Northwestern	2.5	4.1	1.8	1.1	2.3	1.6	2.7	3.5	0.8	1.1
Columbia	3	4.9	2.3	1.3	2.8	1.6	3.1	1	5	2.4
UPenn	2.2	3.5	1.8	0.9	2.4	1	3.1	1.1	0.3	5.1
Others	1.1	3.8	0.5	0.4	1	0.4	1	0.2	0.3	0.6

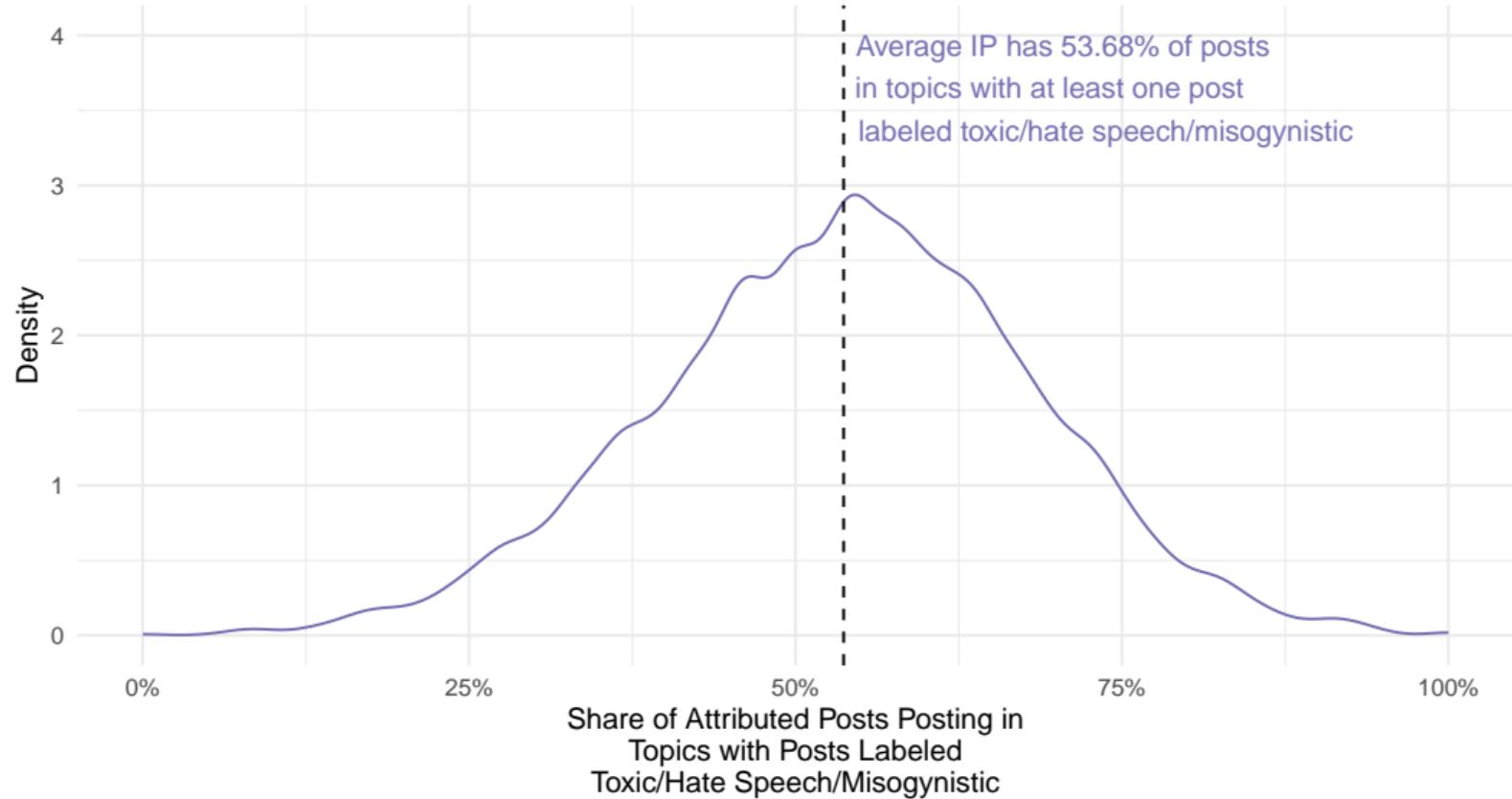
Everybody on EJMR loves talking about themselves ... and about MIT.

Distribution in Toxic Speech across IPs

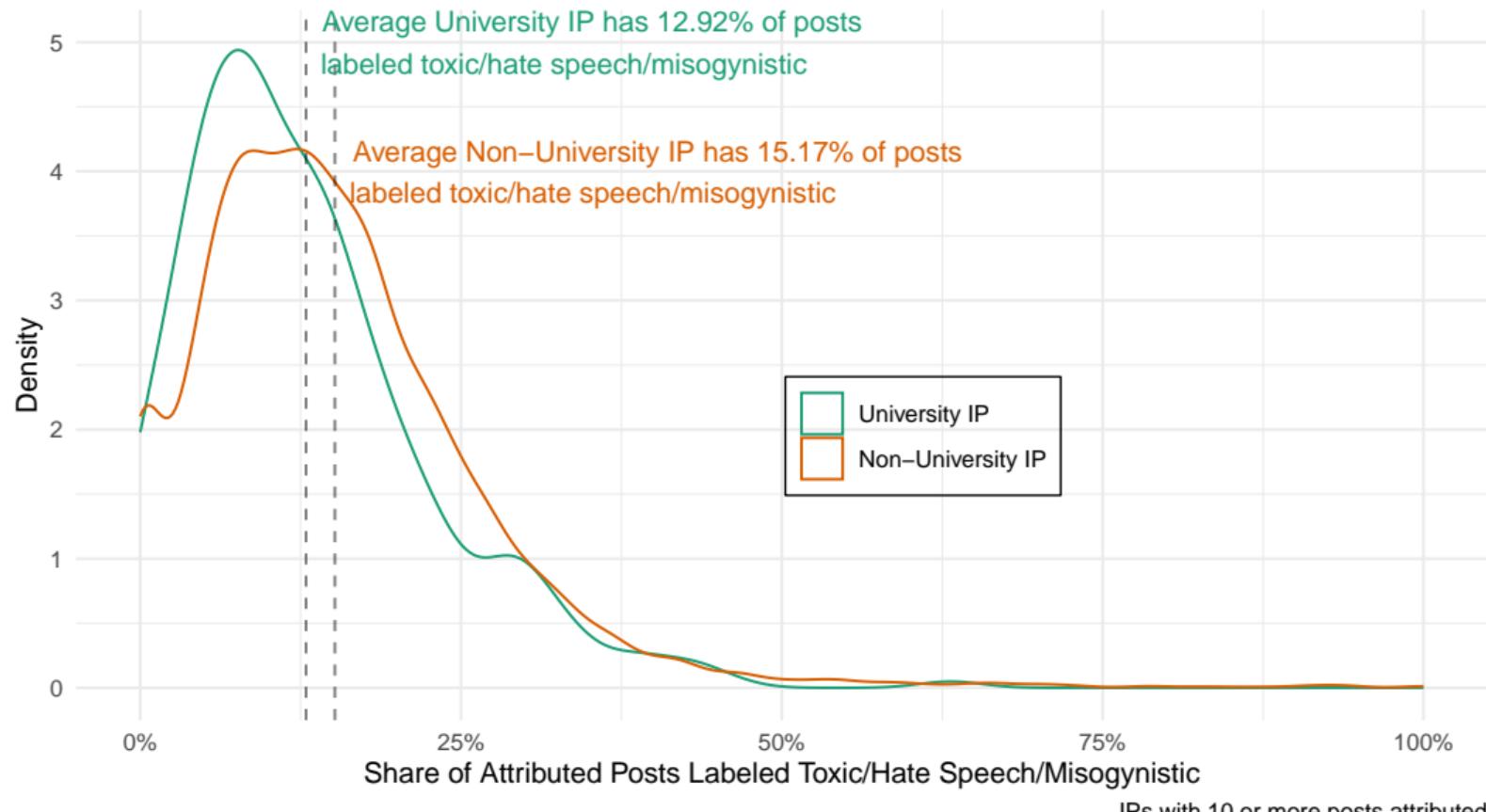


IPs with 10 or more posts attributed

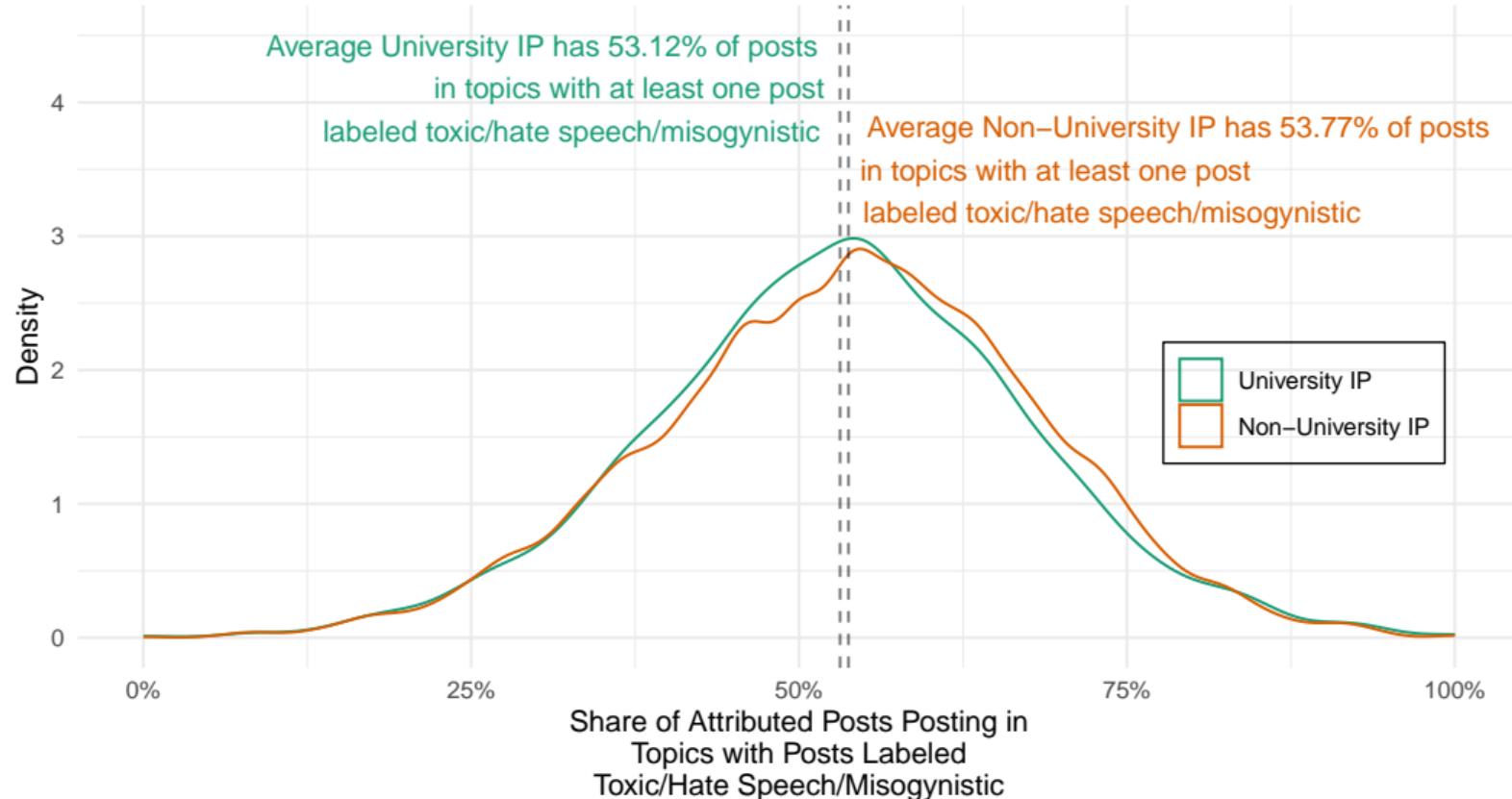
Distribution in Participation in Toxic Conversation



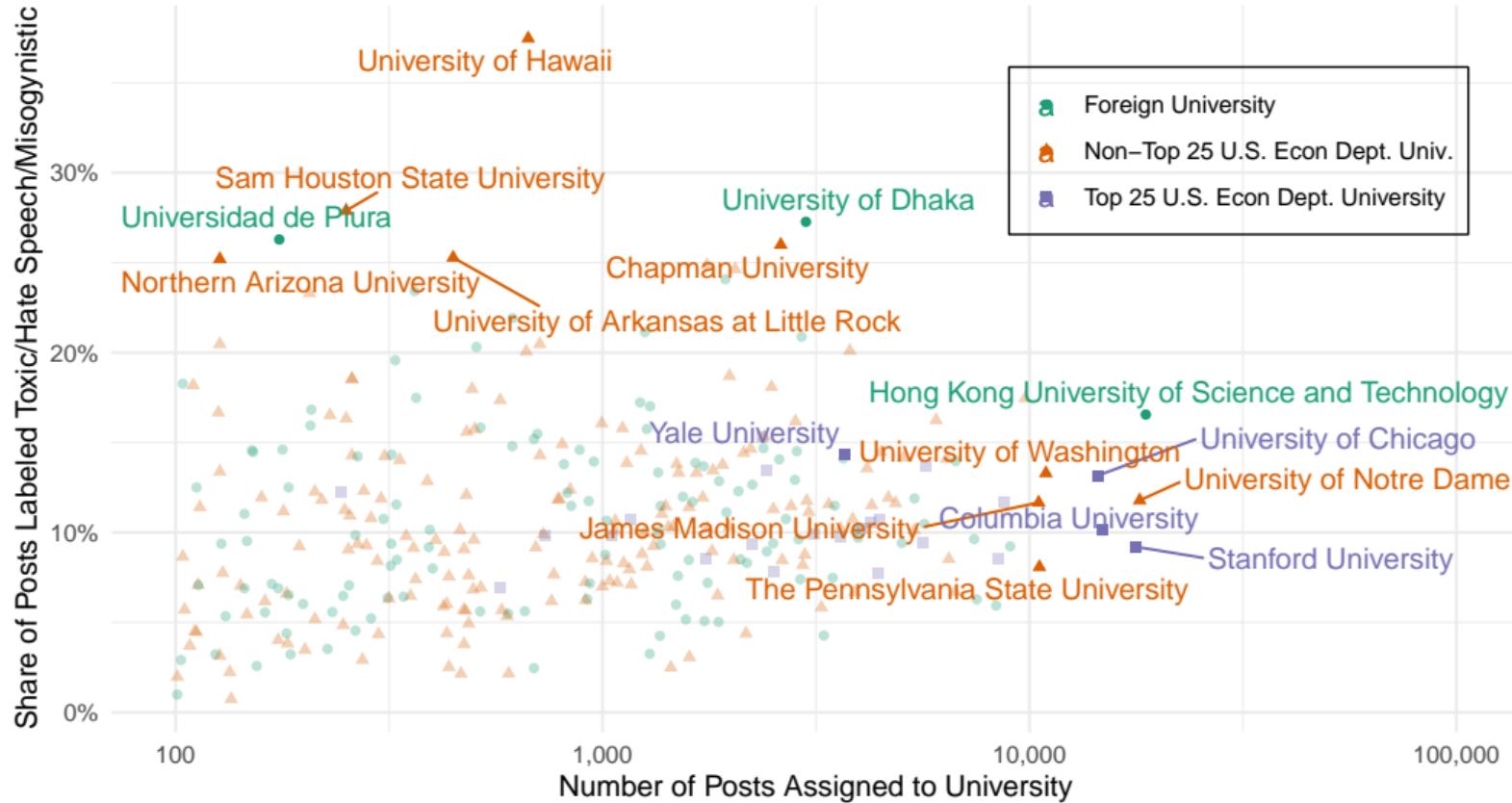
Distribution of Toxic Speech Across IPs by University



Distribution in Participation in Toxic Conversation by University



Toxic Speech on EJMR by University ISPs

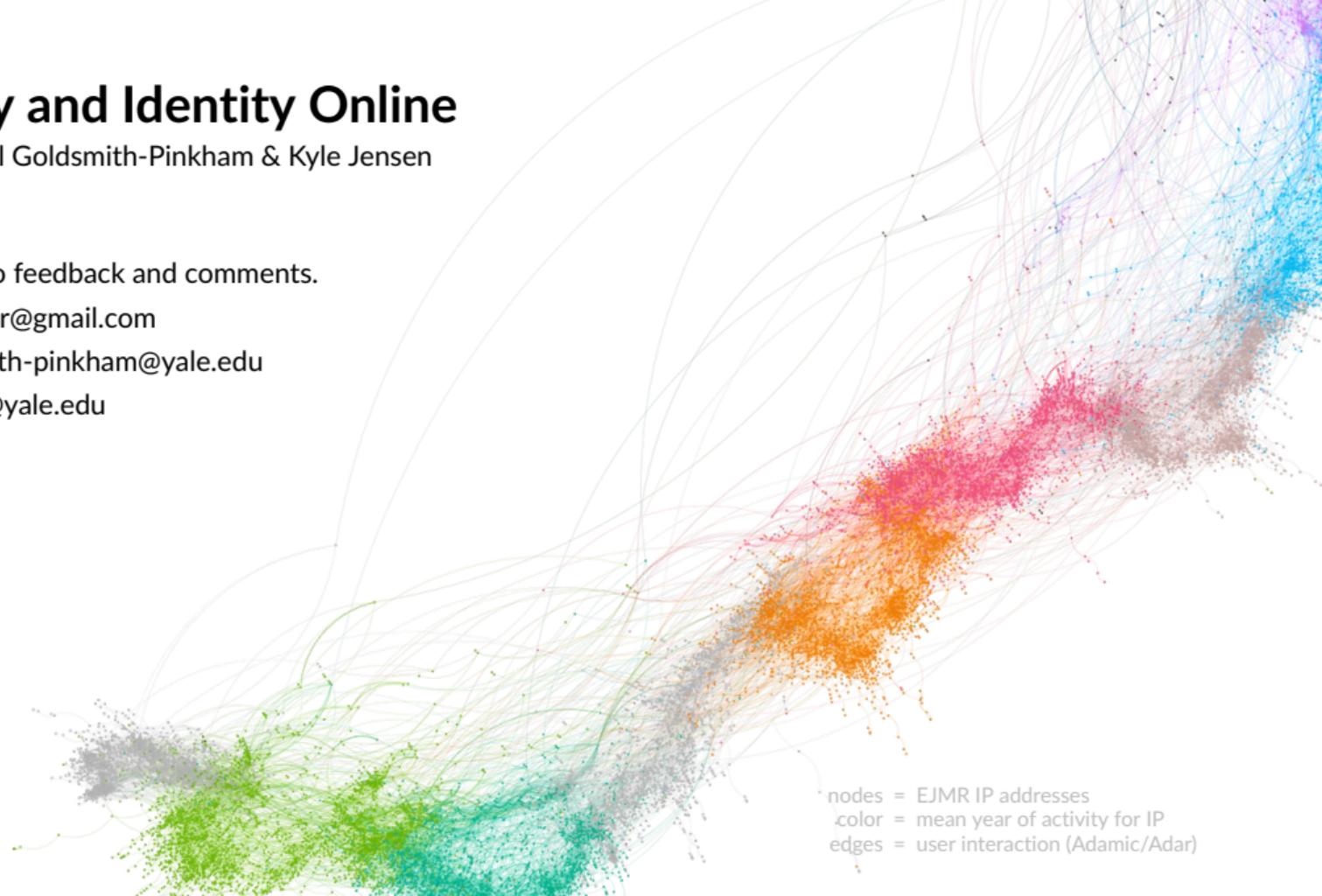


Anonymity and Identity Online

Florian Ederer, Paul Goldsmith-Pinkham & Kyle Jensen

We look forward to feedback and comments.

- florian.ederer@gmail.com
- paul.goldsmith-pinkham@yale.edu
- kyle.jensen@yale.edu

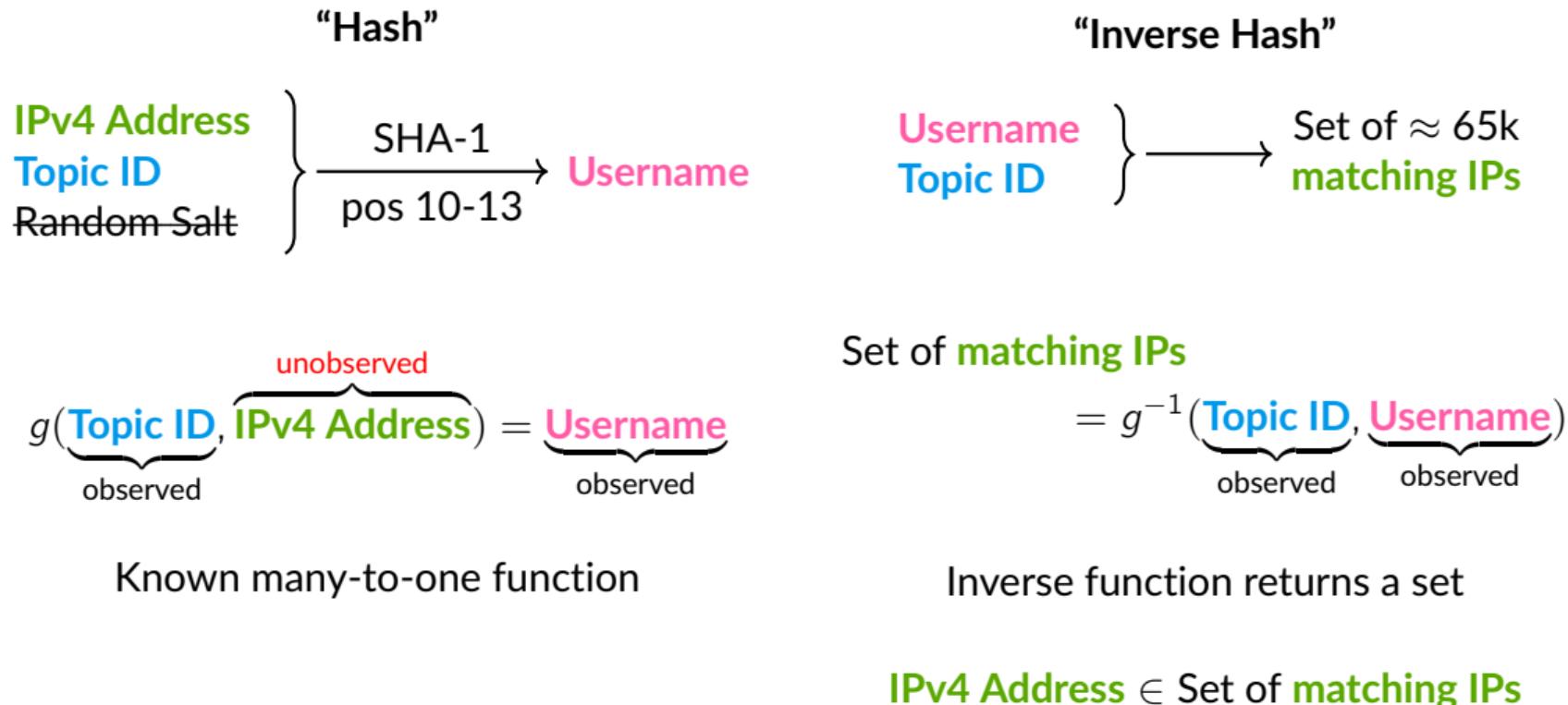


Appendix

From usernames to IP addresses (in practice)

- Hash inversion is conceptually simple but intolerably slow in practice.
 - 695,364 topics \times 2^{32} possible IPv4 addresses \approx 3 quadrillion hashes
 - Check which hashes correspond to observed topic-username combinations
 - Easy to write code in Python, but would take over 60 years on a modern CPU
- Computation is tractable with graphical processing units (GPUs).
 - Basically the same as Bitcoin mining
 - Hash inversion is highly parallelizable
 - 230 hours of total computing time on Nvidia A100 devices each with 6,912 cores
- Hash inversion is only feasible because there is **no salt**.

From usernames and topics to IP addresses



Finding Active IP Addresses

- Statistical properties
 - The noise IPs that match a topic-username (t, u) are **uniformly** distributed across the IPv4 space.
 - The probability that a noise IP a is observed in any particular set $A_{(t,u)}$ containing IPs that match (t, u) is a **hypergeometric** distribution which depends on the number of u in t .
 - The number of times n_a that a noise IP a is observed across all matching IP sets $A_{(t,u)}$ follows a **Poisson binomial** distribution.
- Our approach
 - Null hypothesis that observed counts of an IP a are generated purely by noise
 - Calculate probability under the null hypothesis that an IP a would be observed n_a times *by chance* (i.e., calculate p-values for each a)
 - Rejection of null hypothesis \rightarrow IP a is active

Attributing Active IP Addresses to Posts

- IP address Attribution as an optimization problem
 - Find a small covering set of active IPs for the observed posts.
 - Compute p-values for each a in the set $A_{(t,u)}$ for all (t, u) .
 - For each (t, u) , identify the IP with the lowest p-value.
 - If its p-value $< p^*$, attribute to that IP. Otherwise, leave (t, u) unattributed.
- Simple intuition of this optimization approach
 - 65k IP addresses can explain any given post, but imagine one of these IPs *also* explains many posts in other topics around the same time.
 - What is the likely origin IP of this post? It's probably this highly explanatory IP.

Algorithmic IP address attribution approach

- Key assumptions
 - Sparsity of IP posters + uniformity of hash over full IPv4 space
- Potential issues with current approach
 1. Multiple-hypothesis testing problem / inference on winners
 2. A given IP will show up $N\pi_0$ times randomly, even under the null hypothesis.
- Solutions
 1. Choose conservative p-value thresholds $p^* \approx 10^{-11}$ based on known null distribution
 2. Window the data in relatively short time intervals (7 days, 31 days, 91 days)
- Work in progress
 - Generative model to construct probability statements for each post and IP combination

Choosing the p-value threshold p^*

- Use a wrong substring of SHA-1 hash to construct a pure noise baseline
 - Repeat the entire hash inversion with incorrect hash positions.
 - Compute p-values and attribute post to IP with lowest p-value.
 - Calculate p^* such that we would obtain zero attributions of posts to an IP
- Window-specific p-value thresholds p^*

$$p_{7d}^* = 1.37 \times 10^{-10}$$

$$p_{31d}^* = 2.51 \times 10^{-11}$$

$$p_{91d}^* = 1.39 \times 10^{-11}$$

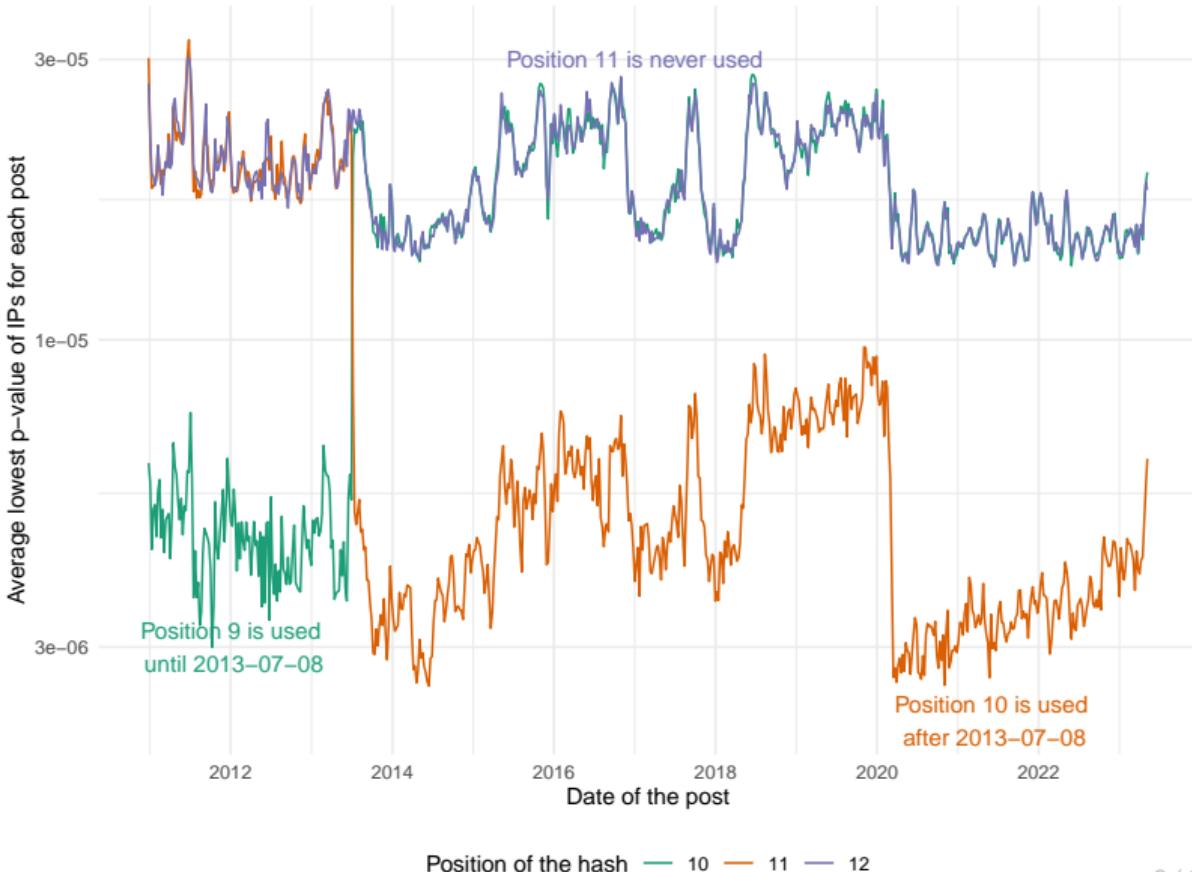
- With these p^* , the number of IP addresses that never posted to EJMR but that we mistakenly attribute to any of the roughly 7 million posts is, in expectation, **less than one**.

Summary of Hash Inversion and IP Attribution Steps

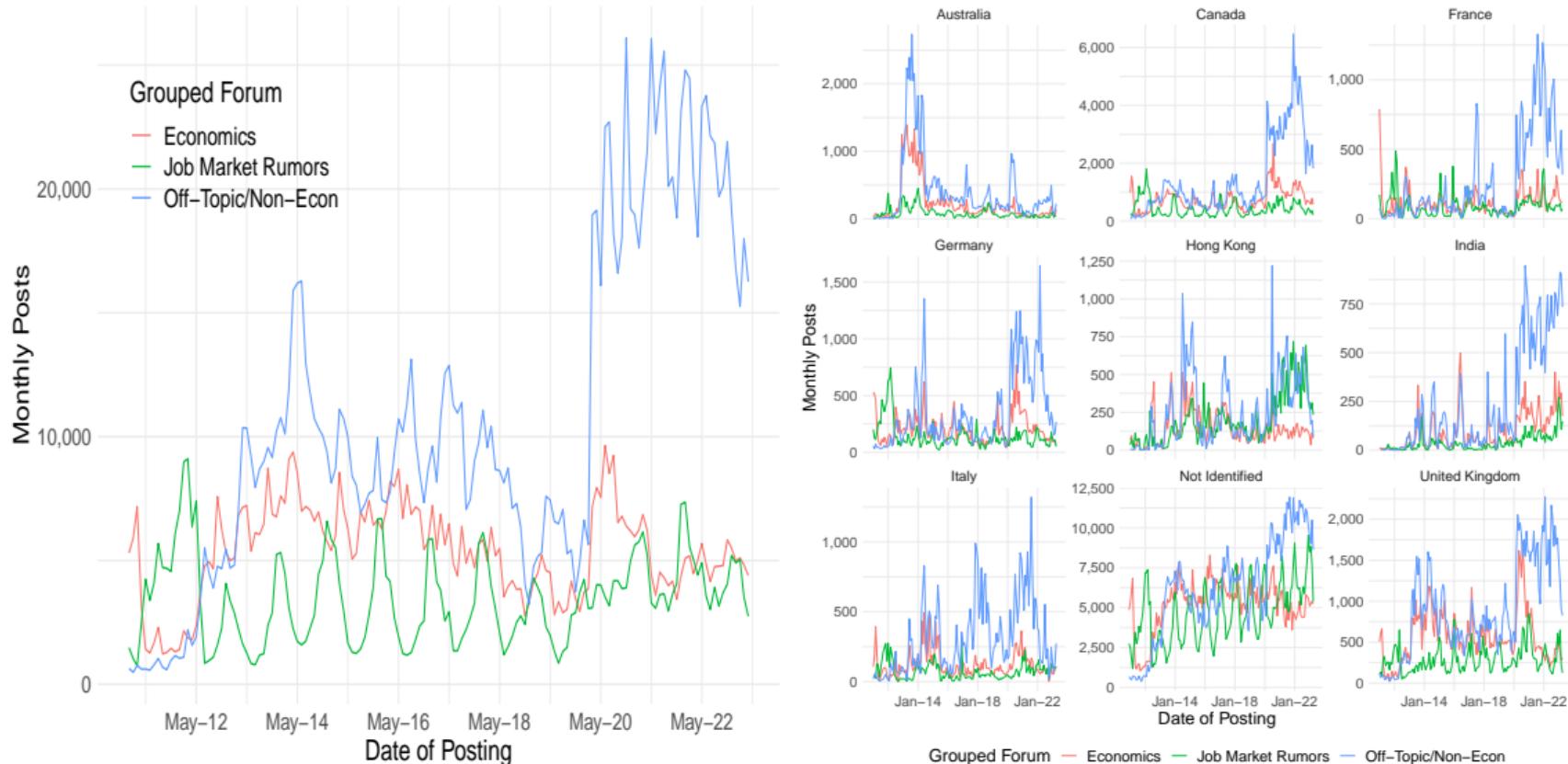
1. Create **topic**-specific **usernames** from hashes of all possible **topic-IP** pairs
2. For each **topic-username** find the set of **matching IPs**
3. Evaluate which **matching IPs** occur “much too often” than expected by random chance (“active IPs”) in a short time window (7 days)
4. Attribute **active matching IP** with lowest p-value $< p^*$ to post or leave post unattributed

Detecting Hash Changes

- Hash changed on July 8, 2013
- Average minimum p-values are much lower for the correct hash
- Average minimum p-values of incorrect hashes closely track each other

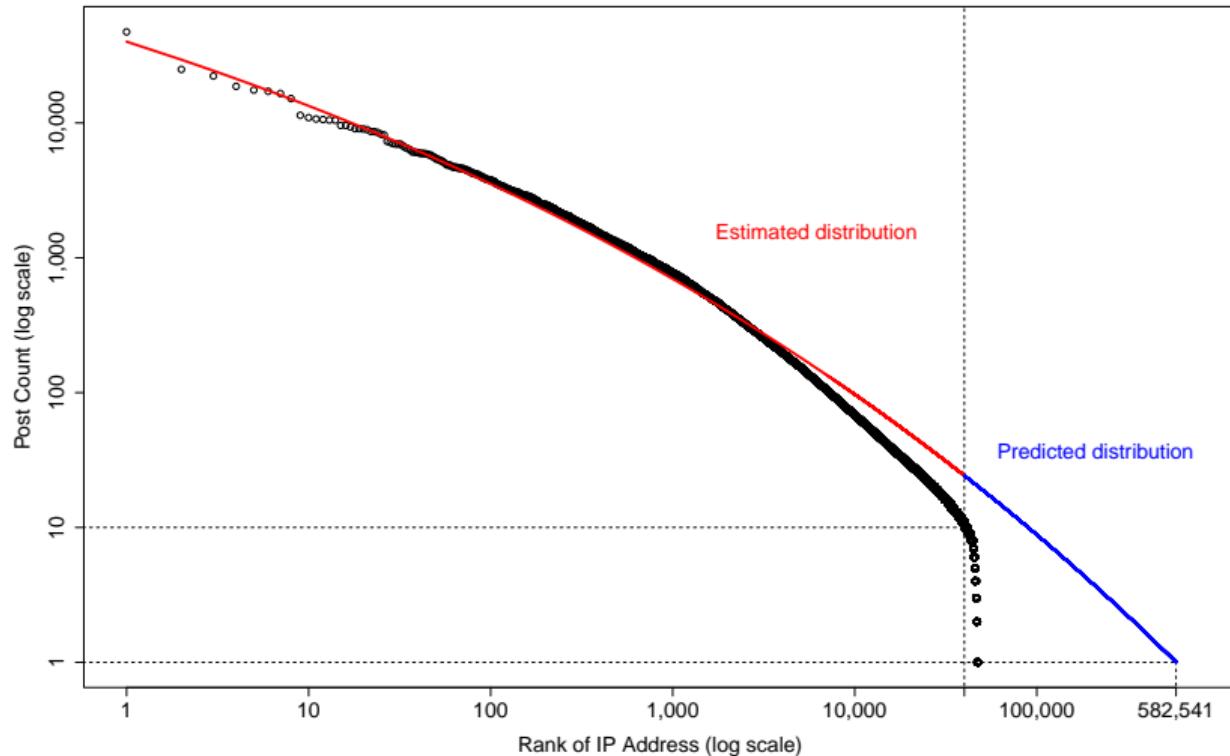


United States and Other Major Countries



Distribution of IP addresses and posts

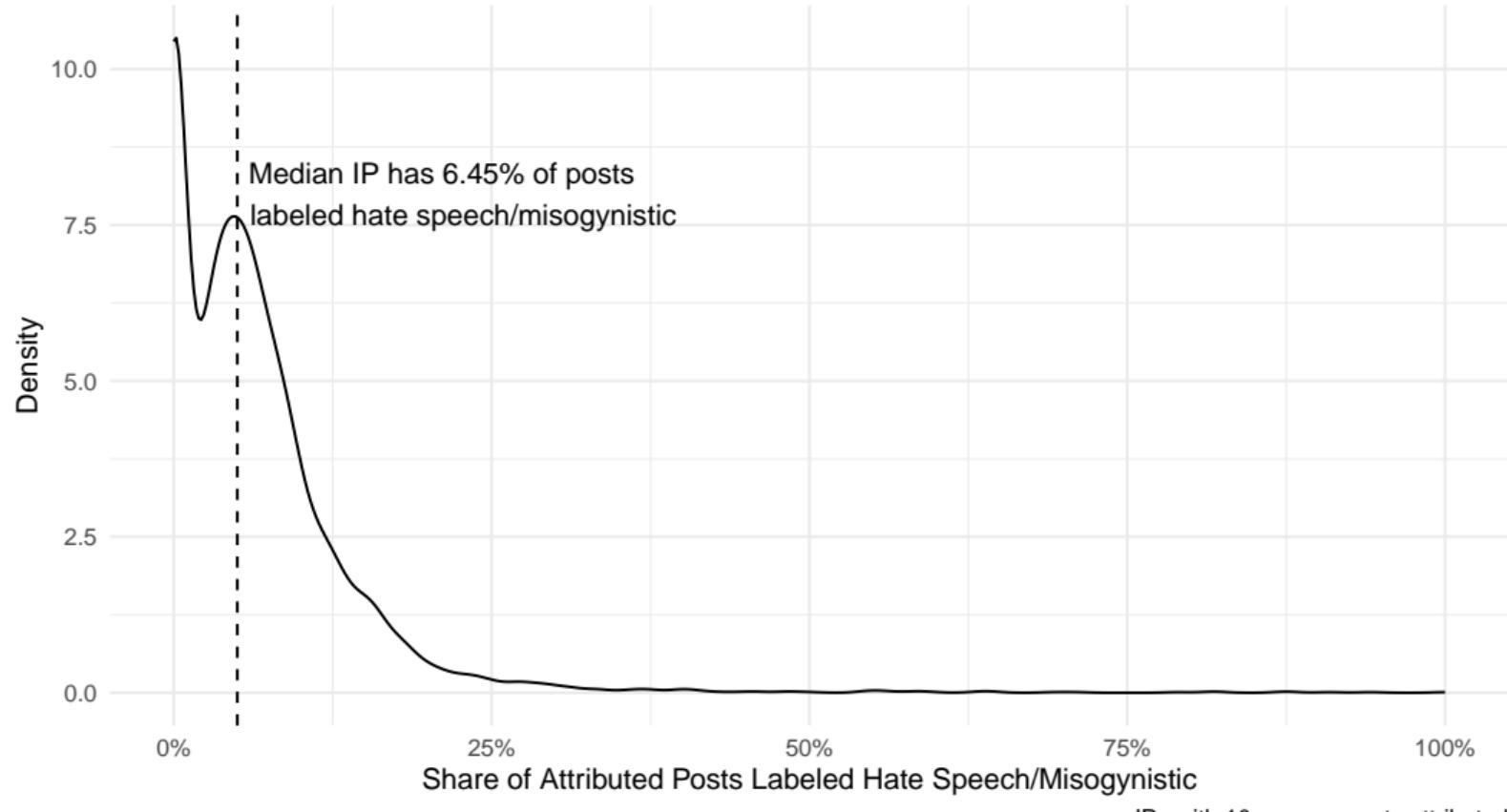
- 66.1% of posts come from 47,630 IPs.
- These **power posters** fit stretched exponential.
- Long tail of **occasional posters** is “unobserved.”
- 582,541 IPs are predicted to have posted on EJMR at least once.



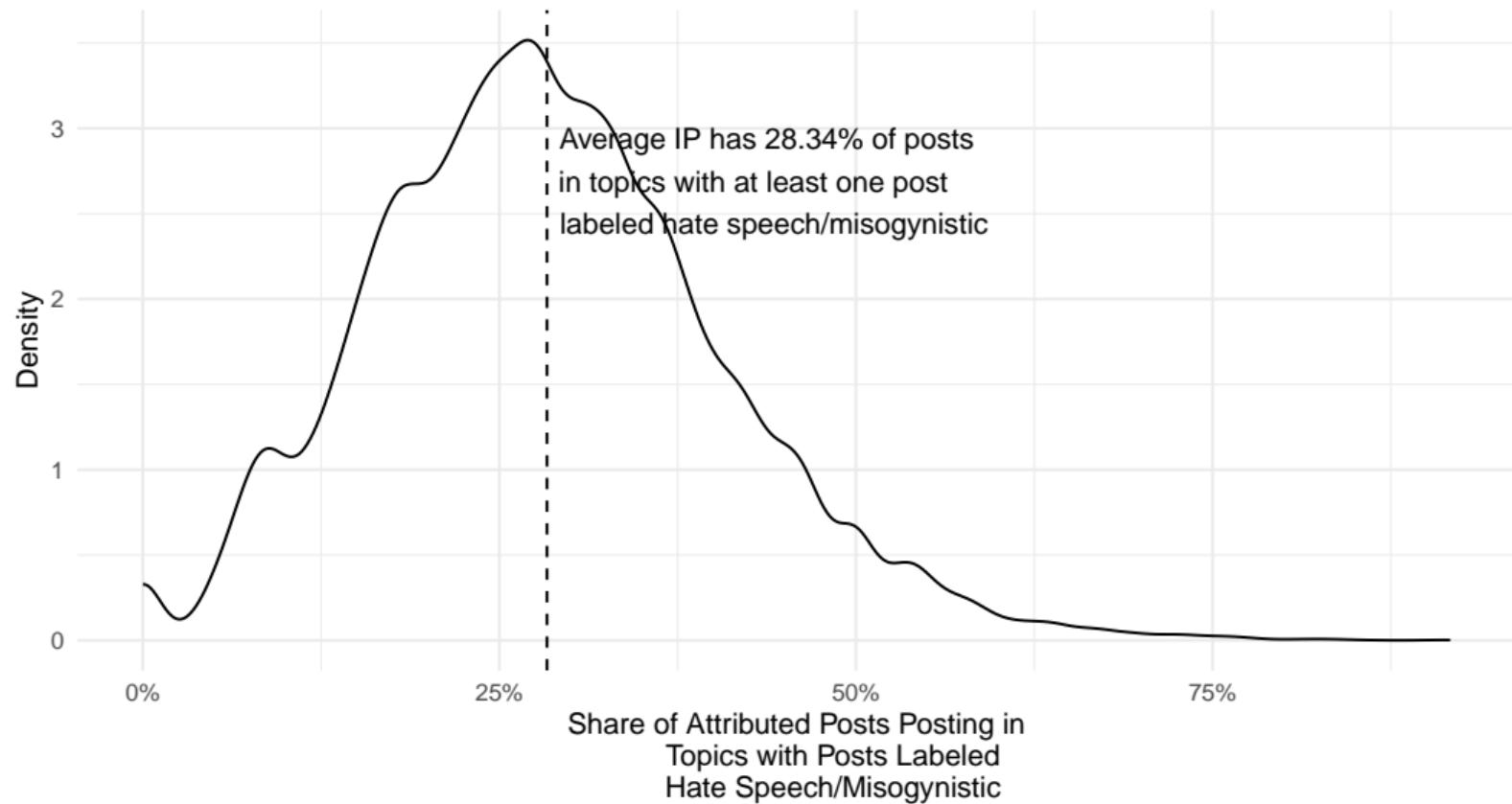
Linguistic Analysis

- EJMR posters use l33tspeak and obfuscation to escape automatic EJMR moderation
 - "Hey a\$\$h01e, I left you a message earlier too. I will be there in Boston to FIEK and RAEP you, so cover your \$hitty a\$\$ and your mouth now." (2014-12-26)
 - "Mold-fa//g//g//ot, I will split your a//s/s in two with my HUMONGOUS super HARD shalong. You will be squealing like the little beia/tch that you are." (2020-01-28)
 - "those d4mn j3ws had no morals either." (2022-08-13)
- Deobfuscation process
 - Collect high-frequency non-English words in English posts to deobfuscate some of the most commonly obfuscated terms
 - Remove common symbol-based obfuscations
 - Transform leetspeak to its canonical form
- Run each post through a number of transformer-based machine learning models for toxicity, sentiment, and misogyny
 - Cross-validate with data by Wu (2020)
 - Note: potential classification error, but unlikely to be correlated across IPs

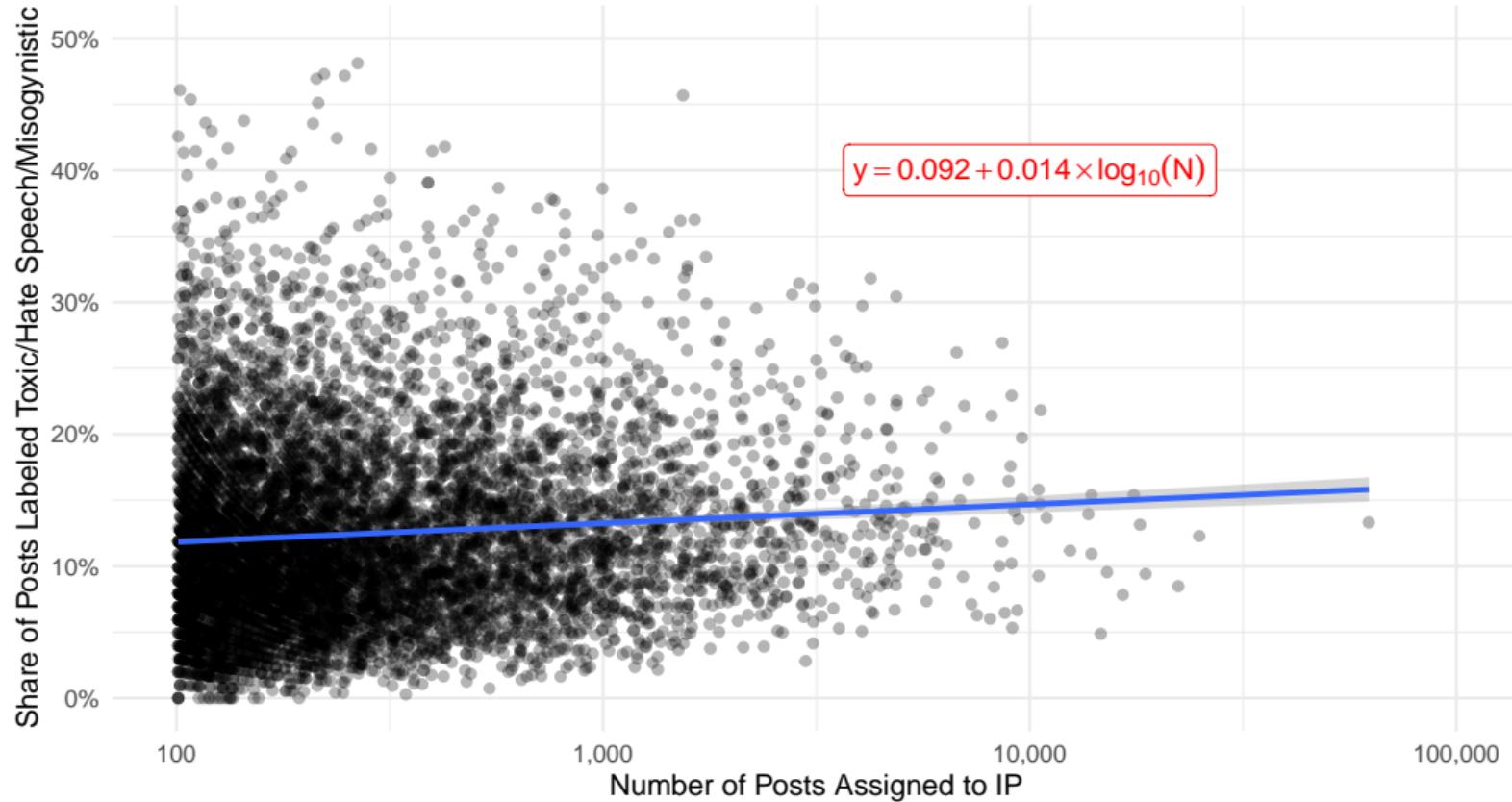
Distribution in Hateful/Misogynistic Speech across IPs



Distribution in Participation in Hateful/Misogynistic Conversation

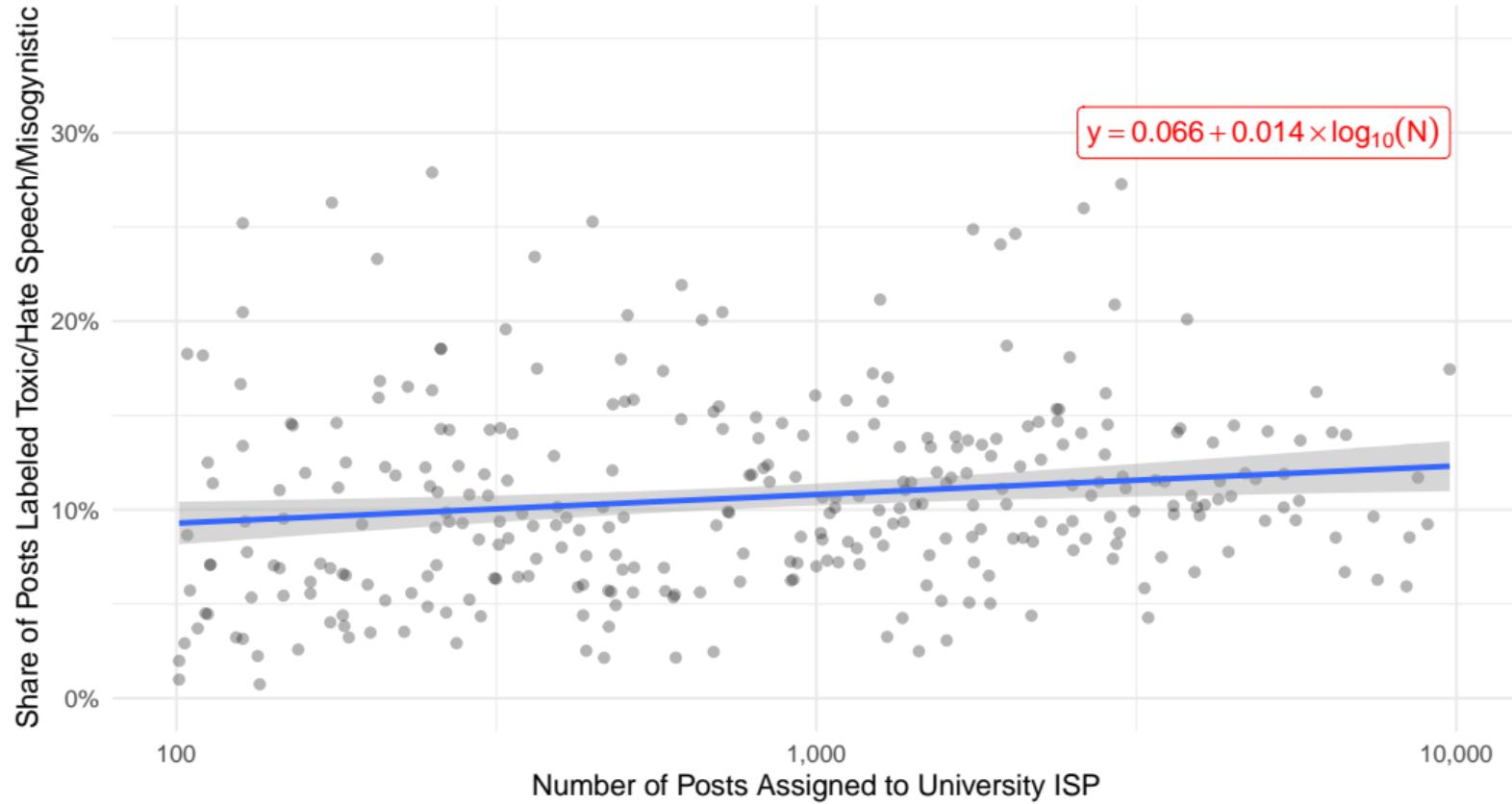


Do frequent IP posters post more toxic speech? Not really.



Note: Only IPs with more than 100 posts are shown.

Toxic Speech on EJMR by University ISPs



Note: Only IPs with more than 100 posts are shown.

References

- Spiegel, H. W.**, "Jacob Viner (1892–1970)," in J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave: A Dictionary of Economics*, Vol. IV, London: Macmillan, 1987, p. 812–14.
- Wu, Alice H.**, "Gender Bias Among Professionals: An Identity-based Interpretation," *Review of Economics and Statistics*, 2020, 102 (5), 867–880.