

Using Data Science to... Study Jerks on the Web

Emiliano De Cristofaro
emilianodc.com





WARNING

CONTENT IN THIS TALK IS
OFFENSIVE
AND UNCENSORED

The Web is weaponized in many ways...





The Web is weaponized in many ways...

Cyberbullying

Cyberaggression

Coordinated
Hate Campaigns

Radicalization

Misogyny



The Web is weaponized in many ways...

Cyberbullying

Cyberaggression

Coordinated
Hate Campaigns

Radicalization

Misogyny

Mis/Disinformation

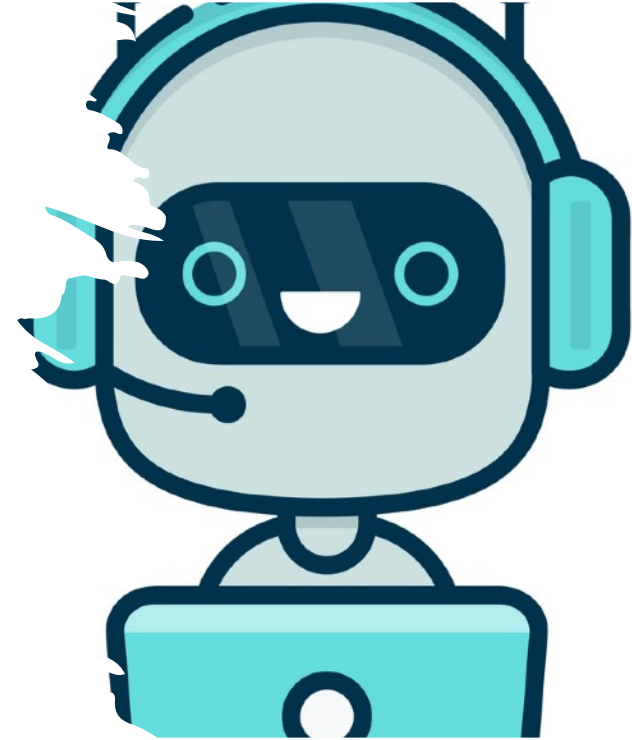
Propaganda

Intentionally Sowing
Discord

Conspiracy Theories

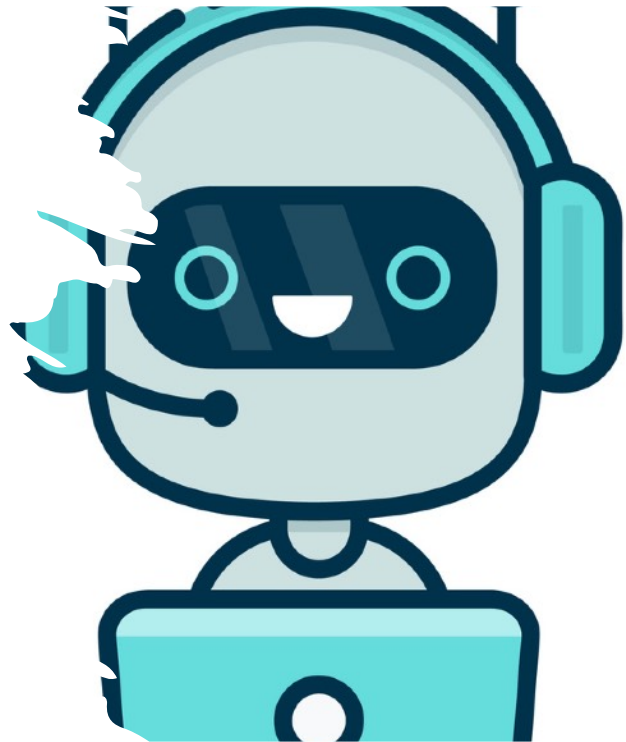
Election Interference

What can we do?



What can we do?

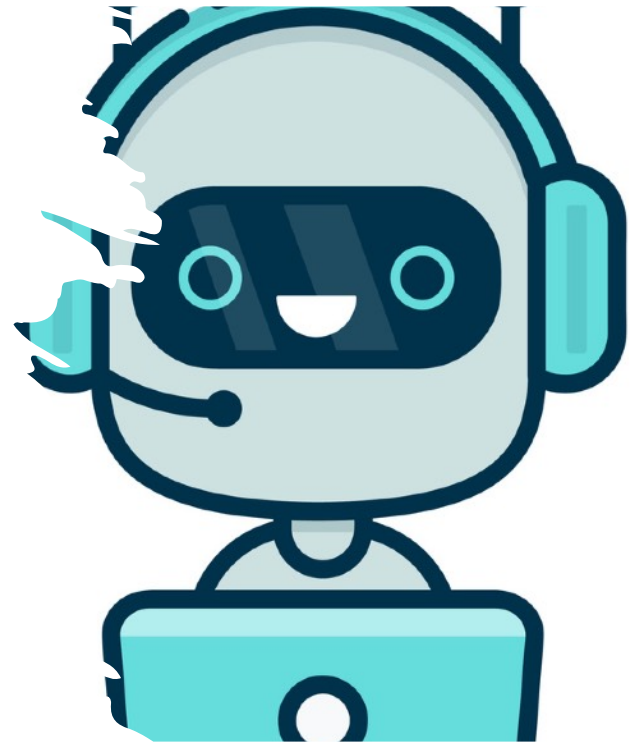
We have experience with **mitigating unwanted activity** on social networks!



What can we do?

We have experience with **mitigating unwanted activity** on social networks!

For instance, we can detect **bot activity** in social media:

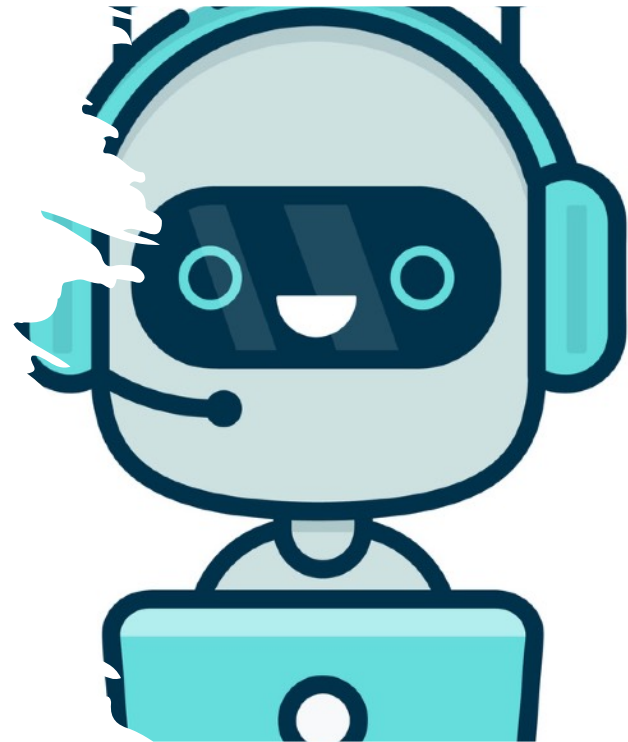


What can we do?

We have experience with **mitigating unwanted activity** on social networks!

For instance, we can detect **bot activity** in social media:

- It's large-scale, synchronized, etc.

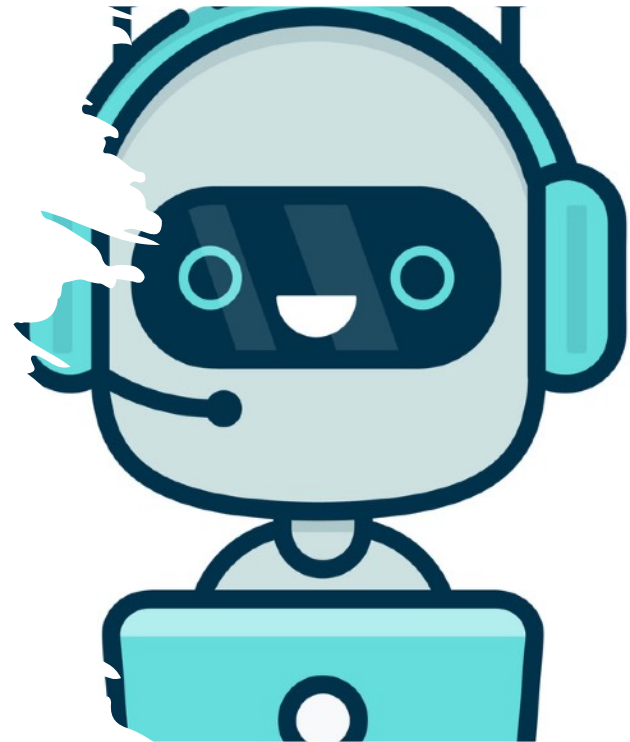


What can we do?

We have experience with **mitigating unwanted activity** on social networks!

For instance, we can detect **bot activity** in social media:

- It's large-scale, synchronized, etc.
- We can build classifiers based on recognizable patterns (SynchroTrap, EvilCohort, Botometer, ...)



What can we do? Really.



What can we do? Really.

What about **human activity**?



What can we do? Really.

What about **human activity**?

- Human activity is **less coordinated**



What can we do? Really.

What about **human activity**?

- Human activity is **less coordinated**
- Characteristic traits stand out less



What can we do? Really.

What about **human activity**?

- Human activity is **less coordinated**
- Characteristic traits stand out less
- We have a **loose synchronization**, common talking points, etc.



What can we do? Really.

What about **human activity**?

- Human activity is **less coordinated**
- Characteristic traits stand out less
- We have a **loose synchronization**, common talking points, etc.

And, we can't take the fight on merely **economic grounds**



What can we do? Really.

What about **human activity**?

- Human activity is **less coordinated**
- Characteristic traits stand out less
- We have a **loose synchronization**, common talking points, etc.

And, we can't take the fight on merely **economic grounds**

- Like for spam, fraud, etc.



What can we do? Really.

What about **human activity**?

- Human activity is **less coordinated**
- Characteristic traits stand out less
- We have a **loose synchronization**, common talking points, etc.

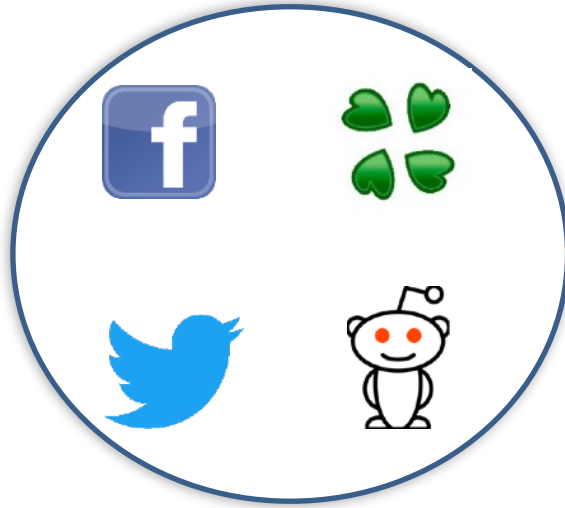
And, we can't take the fight on merely **economic grounds**

- Like for spam, fraud, etc.
- State-level adversaries have deep pockets!

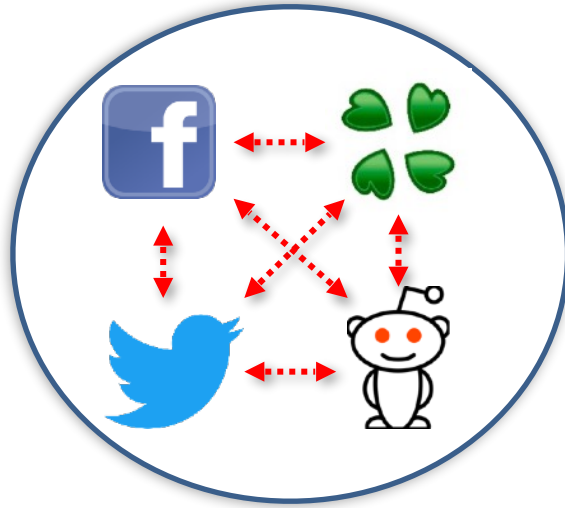


Online services do not exist in a vacuum

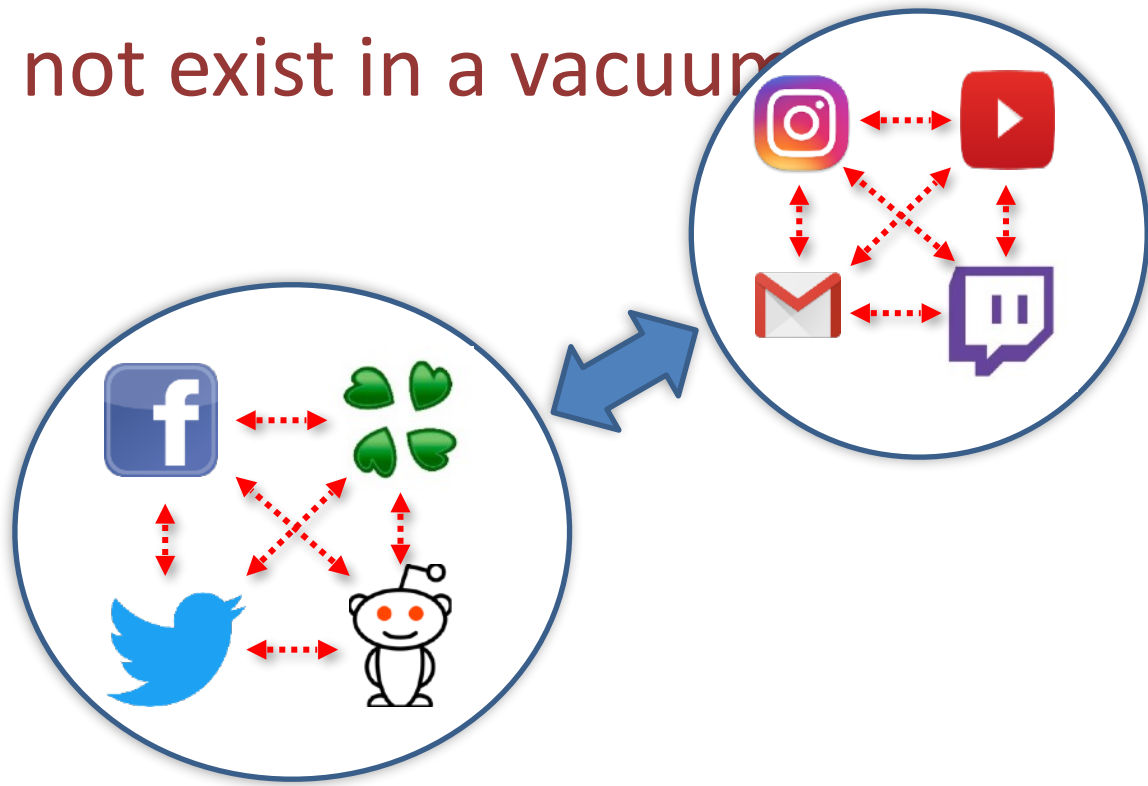
Online services do not exist in a vacuum



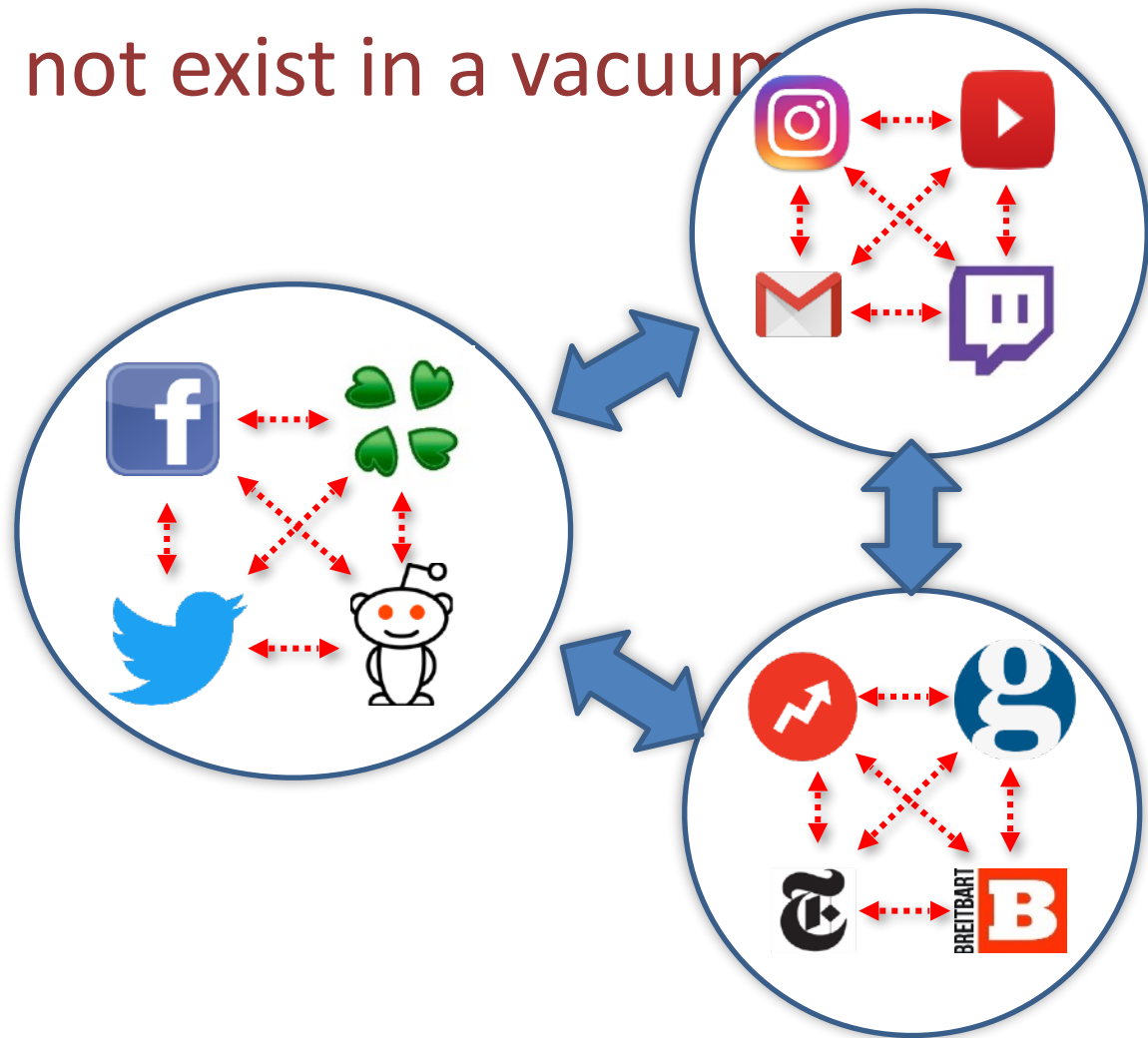
Online services do not exist in a vacuum



Online services do not exist in a vacuum

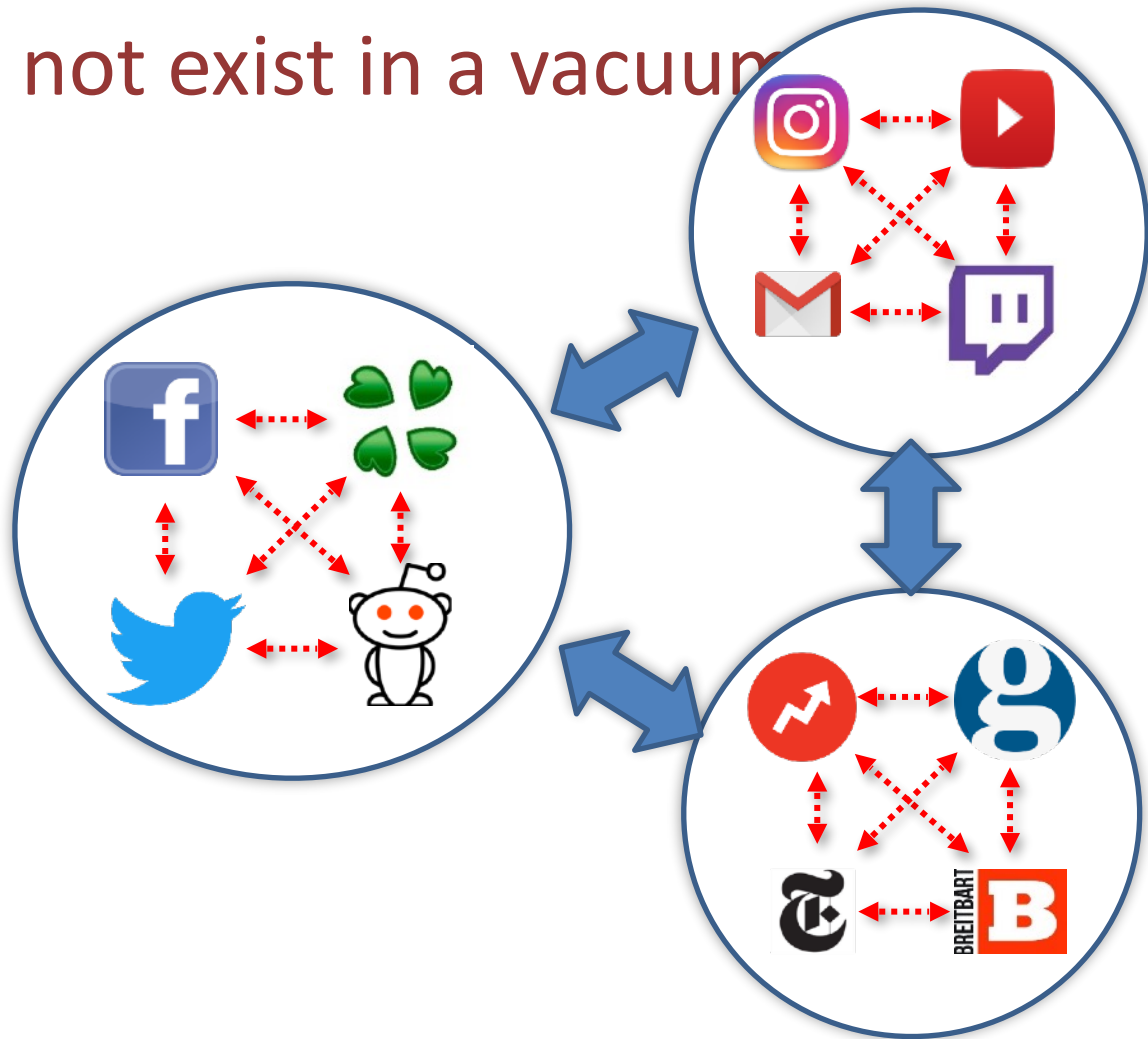


Online services do not exist in a vacuum



Online services do not exist in a vacuum

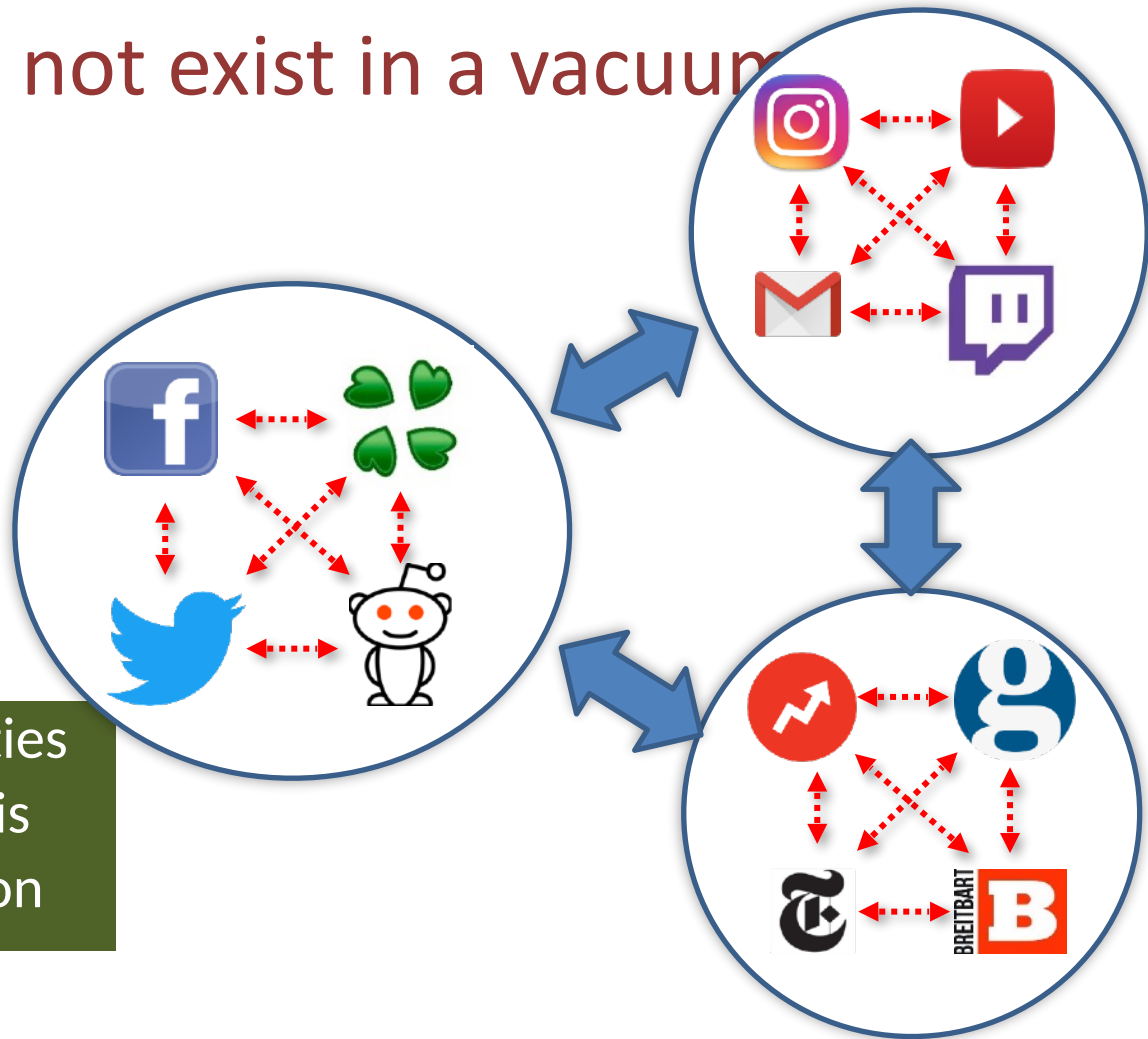
1. Looking at a single service at a time is not enough



Online services do not exist in a vacuum

1. Looking at a **single** service at a time is not enough

2. "Fringe" web communities play a **crucial role** vis-à-vis information weaponization



How a racist, sexist hate mob forced Leslie Jones off Twitter

By Kristen V. Brown



FilmMagic

[f](#) [SHARE](#) [t](#) [TWEET](#) [t](#) [TUMBLR](#) [✉](#) [EMAIL](#)

Over the past 24 hours, Leslie Jones has been inundated with racist, hateful vitriol on Twitter.

How a racist, s mob forced Les off Twitter

By Kristen V. Brown



FilmMagic



Over the past 24 hours, Leslie Jones has been inundated with racist, hateful vitriol on Twitter.

INTERNET

8 DECEMBER 2016

Pizzagate: How a 4Chan conspiracy went mainstream

The power of “meme magic” is changing the world as we know it.

How a racist, s mob forced Les off Twitter

By Kristen V. Brown



FilmM

f SHARE t TWEET t TUMBLR ✉ EMAIL

Over the past 24 hours, Leslie Jones has been inundated with racist, hateful vitriol on Twitter.

INTERNET

8 DECEMBER 2016

Pizzagate: How a 4Chan conspiracy went mainstream

The power of know it. How Does Online Racism Spawn Mass Shooters?

White nationalist terrorism is becoming normalized through internet forums.

BY JAMES PALMER | AUGUST 4, 2019, 6:13 PM



Anti-Muslim Hate Has Been Rampant on Reddit Since the New Zealand Shooting

And the platform doesn't seem to care.



ALI BRELAND

Reporter

[Bio](#)



FilmM

[f](#) [SHARE](#) [t](#) [TWEET](#) [t](#) [TUMBLR](#) [✉](#) [EMAIL](#)

Over the past 24 hours, Leslie Jones has been inundated with racist, hateful vitriol on Twitter.

an conspiracy

Racism Spawn

alized through internet forums.



Anti-Muslim Hate Has Been Rampant on Reddit Since the New Zealand Shooting

And the platform doesn't seem to care.



ALI I
Repo
Bio



SHARE



Over the past 24 hours
hateful vitriol on Twit

CYBERSECURITY

Russia's manipulation of Twitter was far vaster than believed

By **TIM STARKS**, **LAURENS CERULUS** and **MARK SCOTT** | 06/05/2019 06:00 AM EDT | Updated 06/05/2019 08:13 PM EDT



Share on Facebook



Share on Twitter

Russia's infamous troll farm conducted a campaign on Twitter before the 2016 elections that was larger, more coordinated and more effective than previously known, [research](#) from cybersecurity firm Symantec out Wednesday concluded.

an conspiracy

Heard of 4chan?

Heard of 4chan?

He

**I CAN HAS
CHEEZBURGER?**



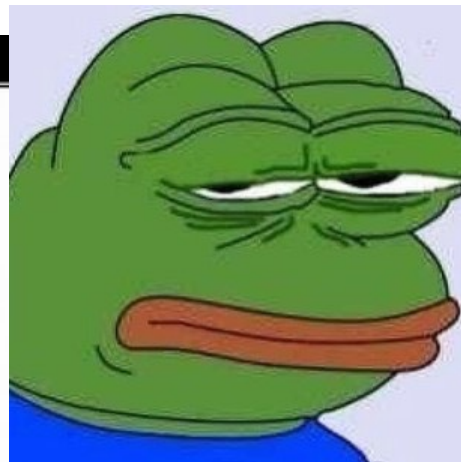
He



He



He



He

I CAN HAS
CHEEZBURGER?



Donald J. Trump
@realDonaldTrump

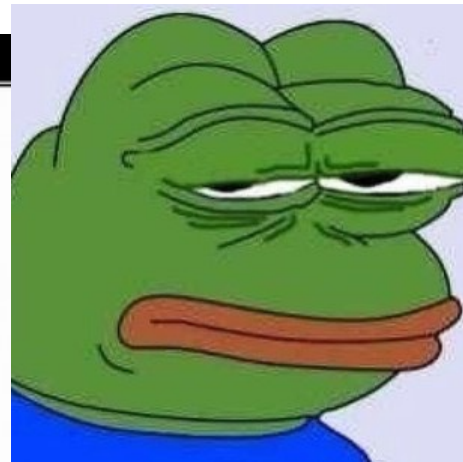
Follow

"@codyave: @drudgereport @BreitbartNews
@WriteinTrump "You Can't Stump the Trump"
[youtube.com/watch?v=MKH6PA...](https://www.youtube.com/watch?v=MKH6PA...) "



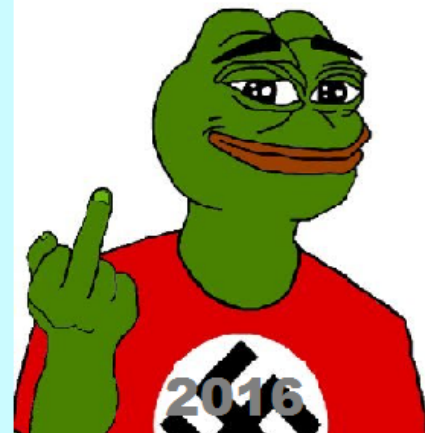
ANONYMOUS

Because none of us are as cruel as all of us.



He

I CAN HAS
CHEEZBURGER?



**WHERE DID EVERYTHING
GO SO REICH**



He



What is 4chan?



What is 4chan?



An image-board forum

Organized in boards (70 at the moment)

What is 4chan?



An image-board forum

Organized in boards (70 at the moment)

An “original poster” (OP) creates a new thread by making a post

Single image attached

What is 4chan?



An image-board forum

Organized in boards (70 at the moment)

An “original poster” (OP) creates a new thread by making a post

Single image attached

Other users can reply:

With or without images, possibly add references to previous posts, quote text, etc.

What is 4chan?



An image-board forum

Organized in boards

An “original post”

Single image attachment

Other users can

With or without

Boards

filter ▼

Japanese Culture

Anime & Manga
Anime/Cute
Anime/Wallpapers
Mecha
Cosplay & EGL
Cute/Male
Flash
Transportation
Otaku Culture
Video Games
Video Games
Video Game Generals
Pokémon
Retro Games

Interests

Comics & Cartoons
Technology
Television & Film
Weapons
Auto
Animals & Nature
Traditional Games
Sports
Alternative Sports
Science & Math
History & Humanities
International
Outdoors
Toys

Creative

Oekaki
Papercraft & Origami
Photography
Food & Cooking
Artwork/Critique
Wallpapers/General
Literature
Music
Fashion
3DCG
Graphic Design
Do-It-Yourself
Worksafe GIF
Quests

Other

Business & Finance
Travel
Fitness
Paranormal
Advice
LGBT
Pony
Current News
Worksafe Requests
Very Important Posts
Misc. (NSFW)
Random
ROBOT9001
Politically Incorrect
International/Random
Cams & Meetups
Shit 4chan Says

Adult (NSFW)

Sexy Beautiful Women
Hardcore
Handsome Men
Hentai
Ecchi
Yuri
Hentai/Alternative
Yaoi
Torrents
High Resolution
Adult GIF
Adult Cartoons
Adult Requests



/pol/ – Politically Incorrect Board

/pol/ – Politically Incorrect Board

File: [sticky.jpg](#) (571 KB, 1600x1131)



☐ **Welcome to /pol/ - Politically Incorrect**

Anonymous (ID: [lv9eTFJC](#)) 🇺🇸 01/08/15(Thu)23:40:44 No.40489590  
[Reply] ▶

This board is for the discussion of news, world events, political issues, and other related topics.

Off-topic and /b/-tier threads will be deleted (and possibly earn you a ban, if you persist). Unless they are quality, well thought out, well written posts, the following are examples of off-topic and/or /b/-tier threads:

- >Red pill me on X. (with no extra content or input of your own)
- >Are X white?
- >Is X degeneracy?
- >How come X girls love Y guys so much?
- >If X is true, then how come Y? Checkmate Z.

The variety of threads allowed here are very flexible and we believe in freedom of speech, but we expect a high level of discourse befitting of the board. Attempts to disrupt the board will not be tolerated, nor will calls to disrupt other boards and sites.

/pol/ – Politically Incorrect Board

File: [sticky.jpg](#) (571 KB, 1600x1131)



Welcome to /pol/ – Politically Incorrect

Anonymous (ID: lv9) 01/01/2023 23:40:44 No.40489590

This
issue

s, political

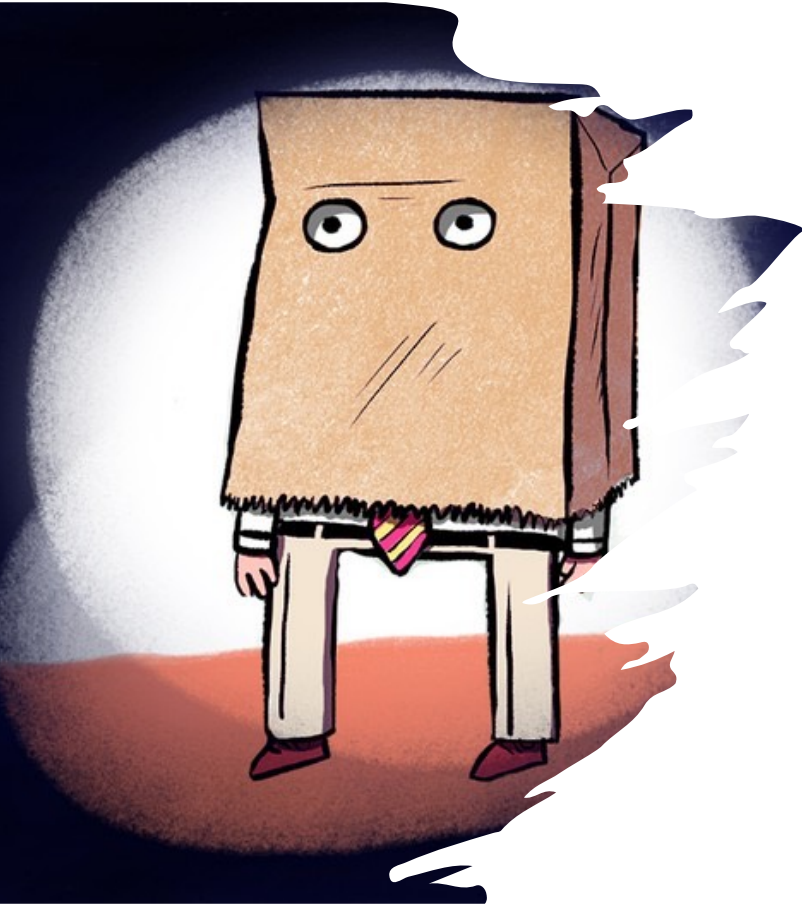
Extremely lax moderation

Almost anything goes

- >Red pill me on X. (with no extra
- >Are X white?
- >Is X degeneracy?
- >How come X girls love Y guys so much?
- >If X is true, then how come Y? Check

The variety of threads allowed here are very flexible. We believe in freedom of speech, but we expect a high level of discourse befitting of the board. Attempts to disrupt the board will not be tolerated, nor will calls to disrupt other boards and sites.

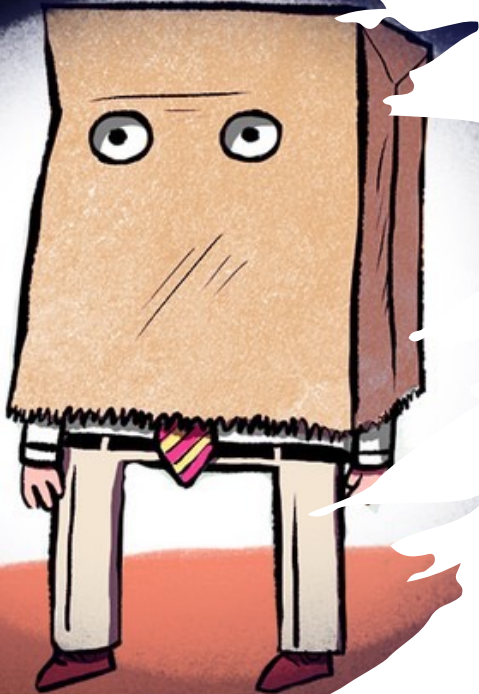
Anonymity & Ephemerality



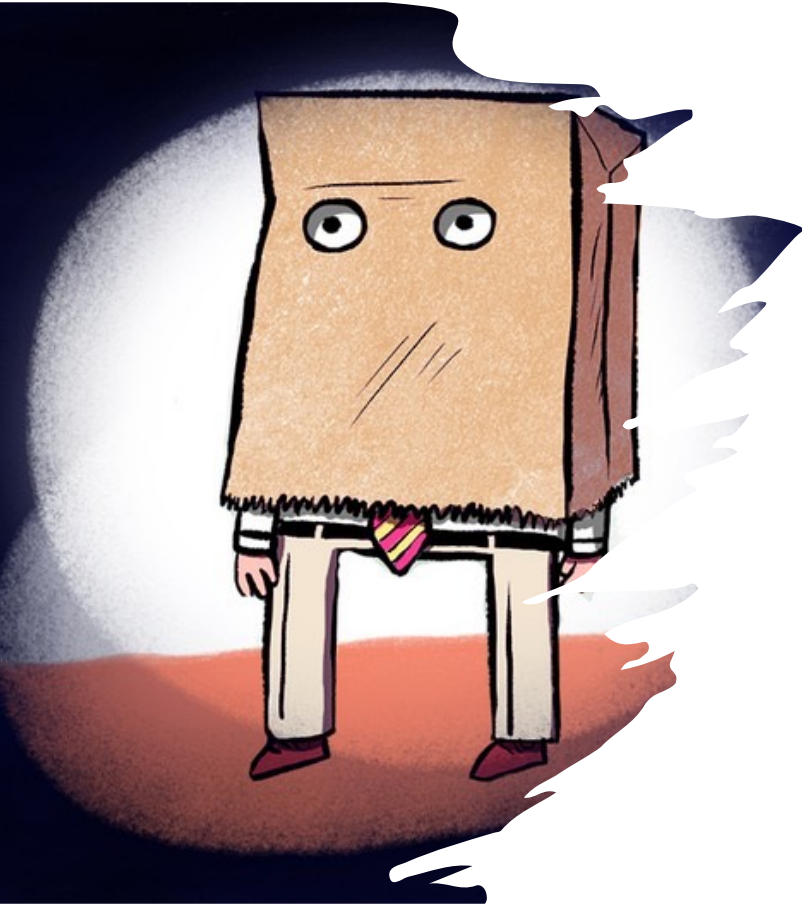
Anonymity & Ephemerality

Users do not need to register an account

Anonymity is the default (and preferred) behavior



Anonymity & Ephemerality

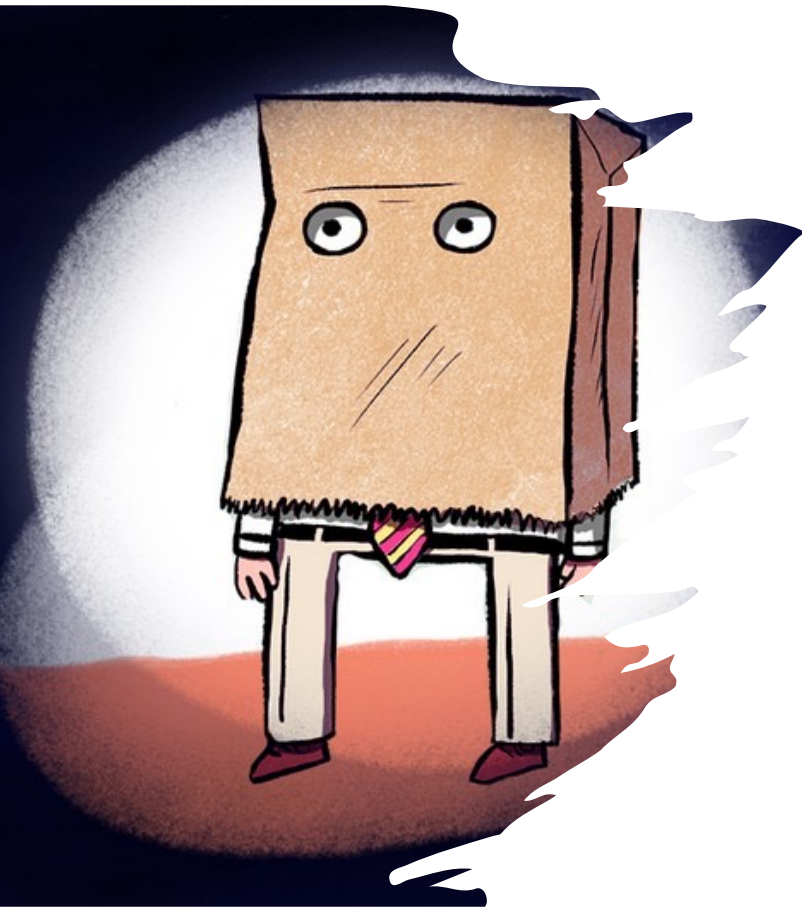


Users do not need to register an account

Anonymity is the default (and preferred) behavior

“Some” degree of permanence and identifiability is supported

Can enter a name along with their posts (no authentication though)



Anonymity & Ephemerality

Users do not need to register an account

Anonymity is the default (and preferred) behavior

“Some” degree of permanence and identifiability is supported

Can enter a name along with their posts (no authentication though)

Threads get “archived” after a while

Actually all posts deleted after a week

Our dataset

A single newline-delimited JSON file

3.5 years worth of data

June 2016 - November 2019

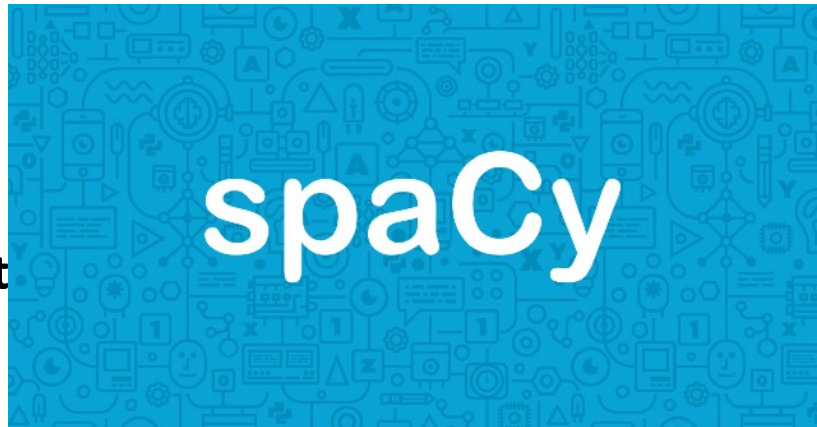
~3.4 million threads from /pol/

Over 134 million posts

Augmented Data with:

Google's Perspective API

spaCy Named Entities mentioned in each post



Hate Speech?



Crowdsourced dictionary

Manually filtered a bit

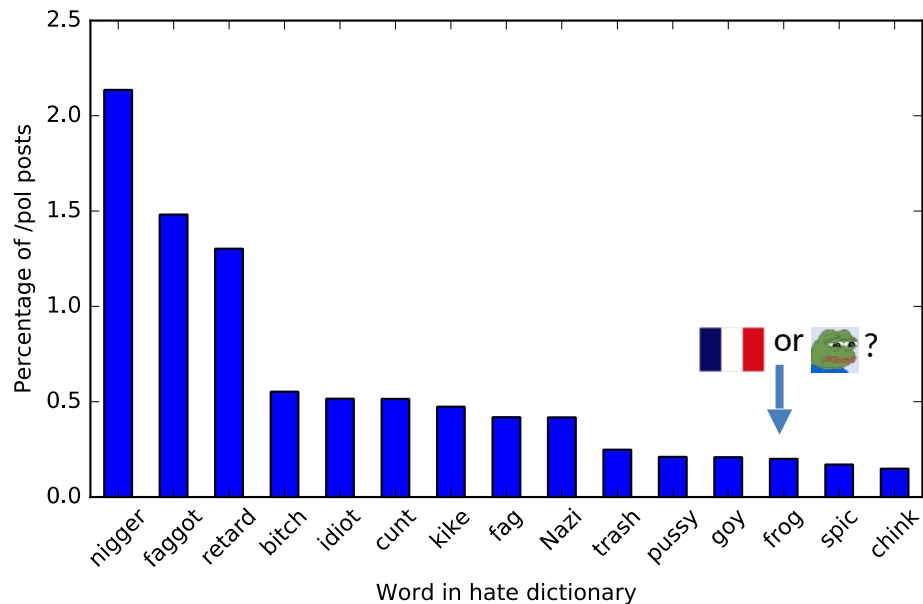
/pol/ by far most
hate speech use

/pol/ 12%

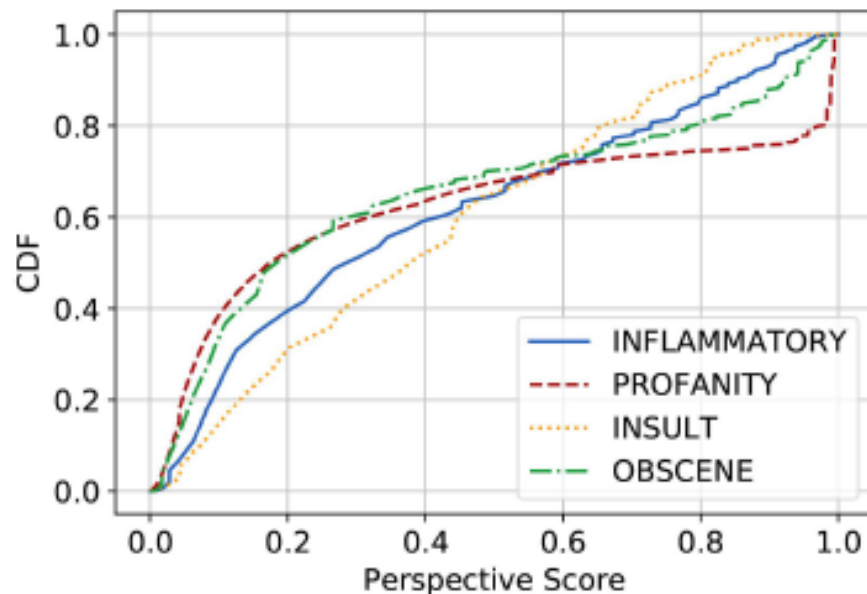
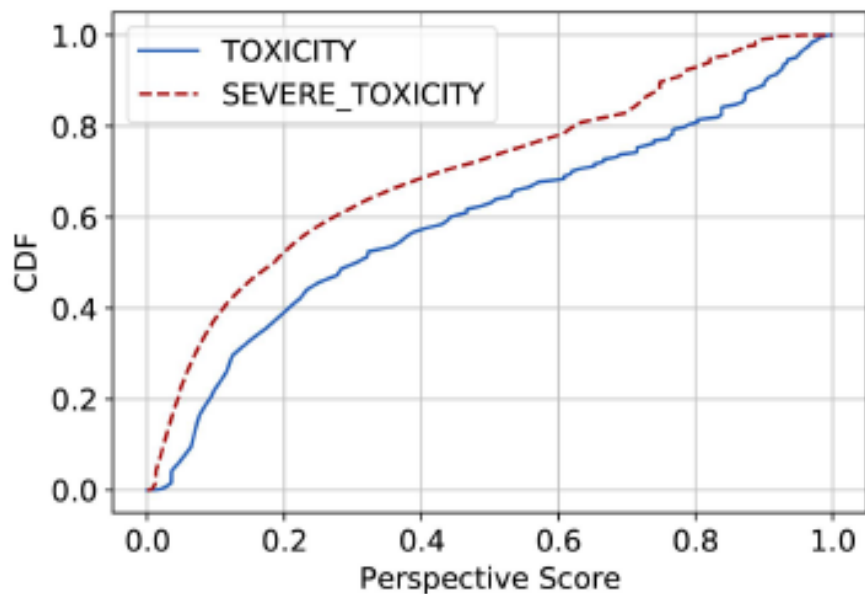
/sp/ 7.3%

/int/ 6.3%

Twitter 2.2%



Google's Perspective API



Challenges of studying fringe communities

Challenges of studying fringe communities

0. Actions not limited to the platform

Challenges of studying fringe communities

0. Actions not limited to the platform

1. Not your typical social networks (e.g., anonymous/ephemeral)

Challenges of studying fringe communities

- 0. Actions not limited to the platform
- 1. Not your typical social networks (e.g., anonymous/ephemeral)
- 2. Knowing what they're talking about is not easy

Challenges of studying fringe communities

- 0. Actions not limited to the platform
- 1. Not your typical social networks (e.g., anonymous/ephemeral)
- 2. Knowing what they're talking about is not easy
- 3. You might get attacked, doxxed, etc.

Coordinated hate campaigns (“raids”)



Coordinated hate campaigns (“raids”)

Someone posts a YouTube link

Maybe with a prompt like “you know what to do”

YouTube video starts getting hateful comments



Coordinated hate campaigns (“raids”)

Someone posts a YouTube link

Maybe with a prompt like “you know what to do”

YouTube video starts getting hateful comments

If raid is taking place:

Peak in YouTube comments while thread alive?

/pol/ thread and YT comments
synchronized?



Raids on other platforms (e.g., YouTube)

Raids on other platform



Hans2k4ever 1 month ago

When will the normies learn?

Reply • 181

[View all 3 replies](#)



Jack Harper 1 month ago

mfw he accidentally redpills his whole class, they all found the memes funny even the so called "hate speech" he talks about

Reply • 13



Jack Harper 1 month ago

I guess I haven't tightened my fedora enough to be as enlightened by my own enlightenment as you sir. Carry on brave knight!

Reply •



Adem Adiyaman 1 month ago

23:21 "the victory of donald trump and the rise of ultra nazi people" you dense fuck, nice brain washing

Reply • 227

[View all 16 replies](#)



Szgerle 1 month ago

Not even just your regular nazi but ULTRA NAZI!

Reply • 1



Drakewood 1 month ago

+Potato Jenova A FUCKING LEAF!

Reply •



Thought Criminal 1 month ago

Holy fuck, (((they))) are actually intimidated by the powers of Kek.

Reply • 169



A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web



Emiliano De Cristofaro

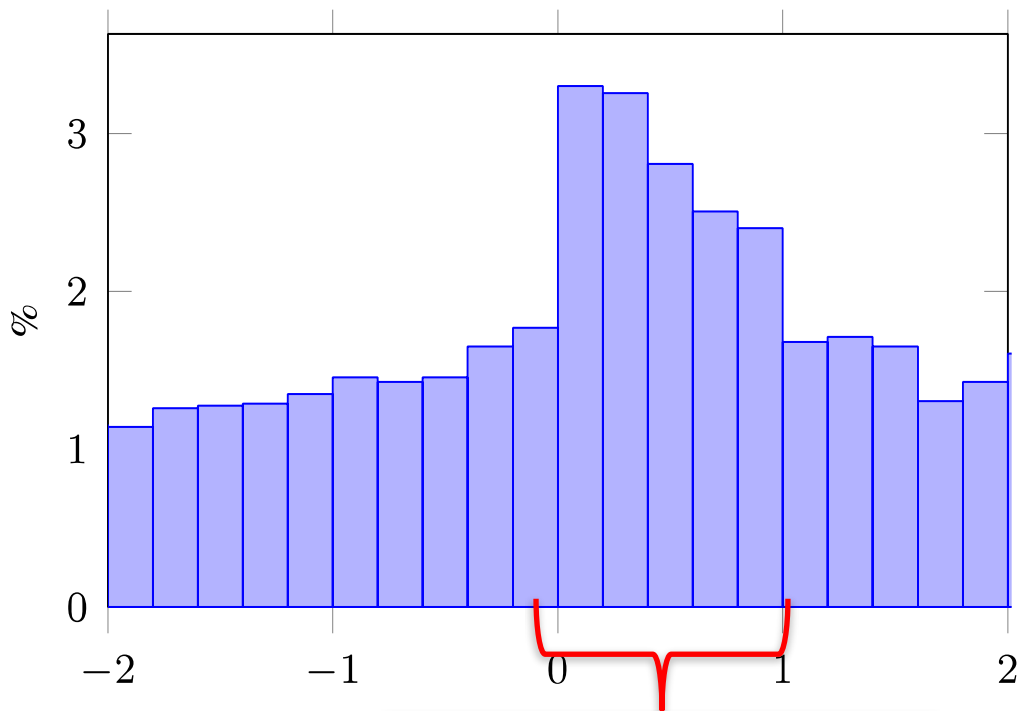
[Subscribe](#) 3

11,013 views

[Add to](#) [Share](#) [More](#)

[Like](#) 80 [Dislike](#) 526

Activity Peaks



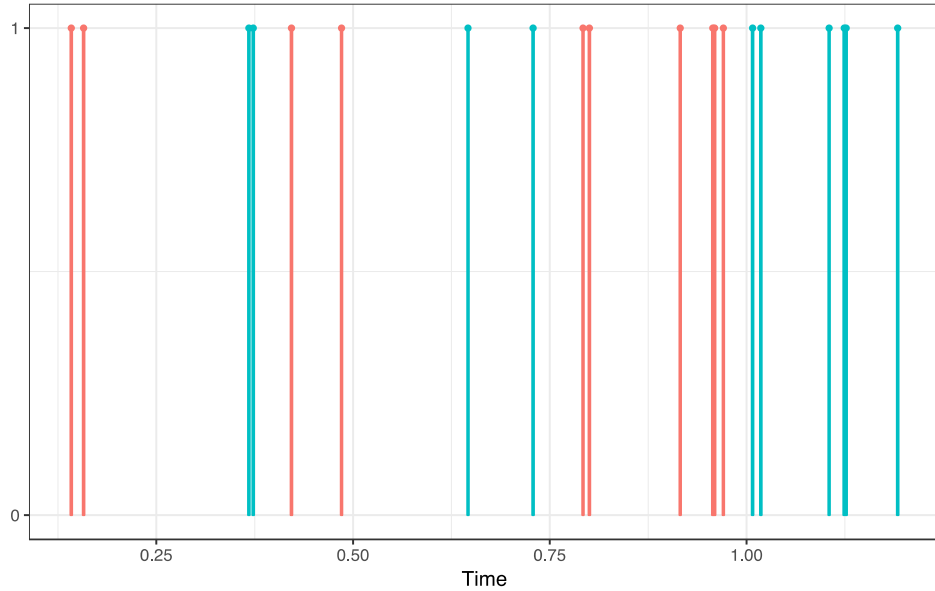
YT videos with peaks during
4chan thread

Determined via PDF of
commenting timeseries

14% of videos see peak
commenting activity during /
pol/ thread lifetime

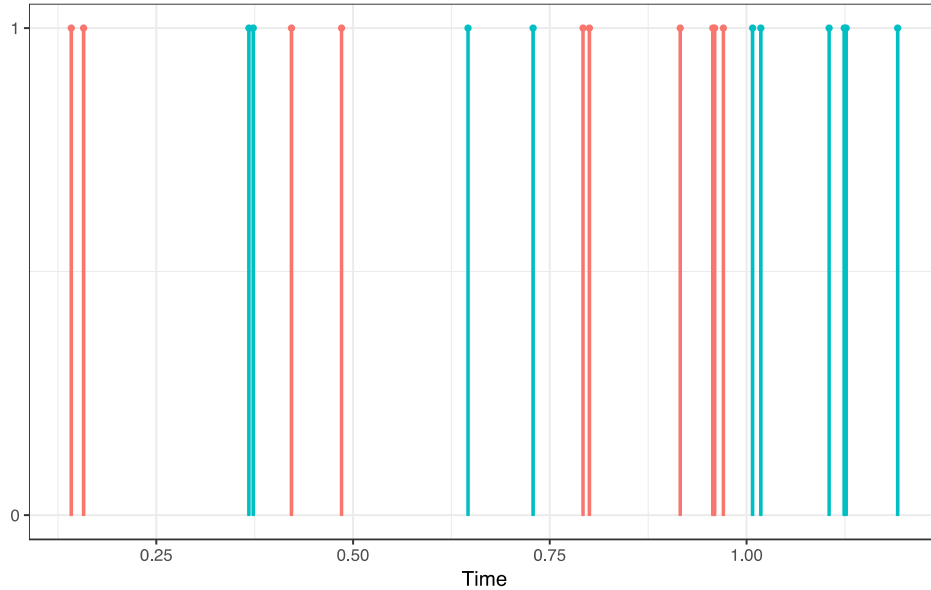
Synchronization

Synchronization

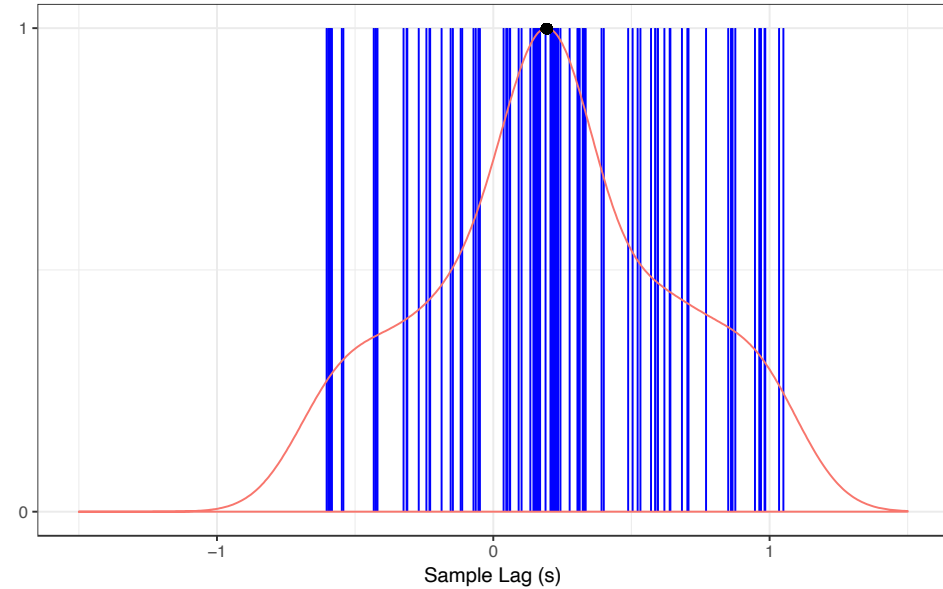


Two series, second randomly shifted from first
by 0.2s on avg

Synchronization

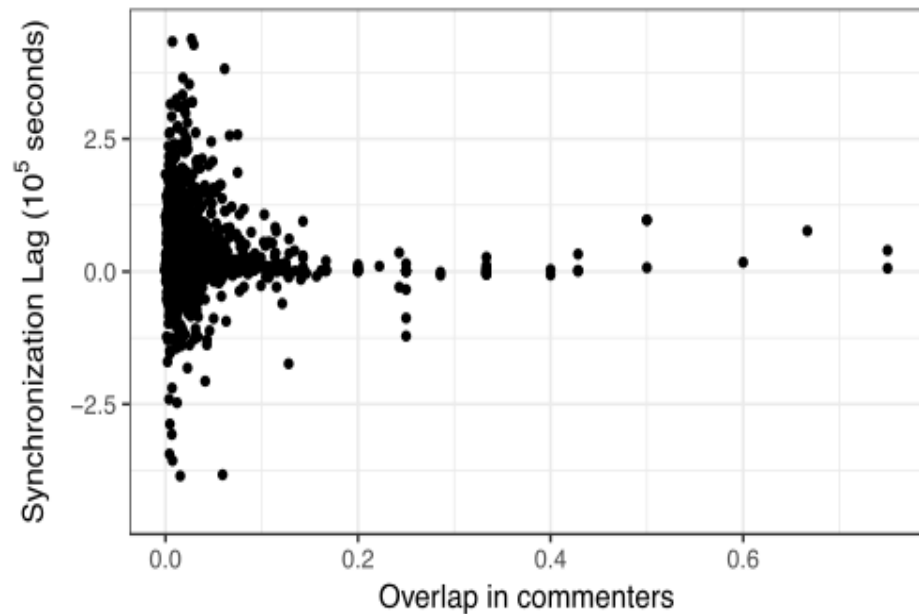
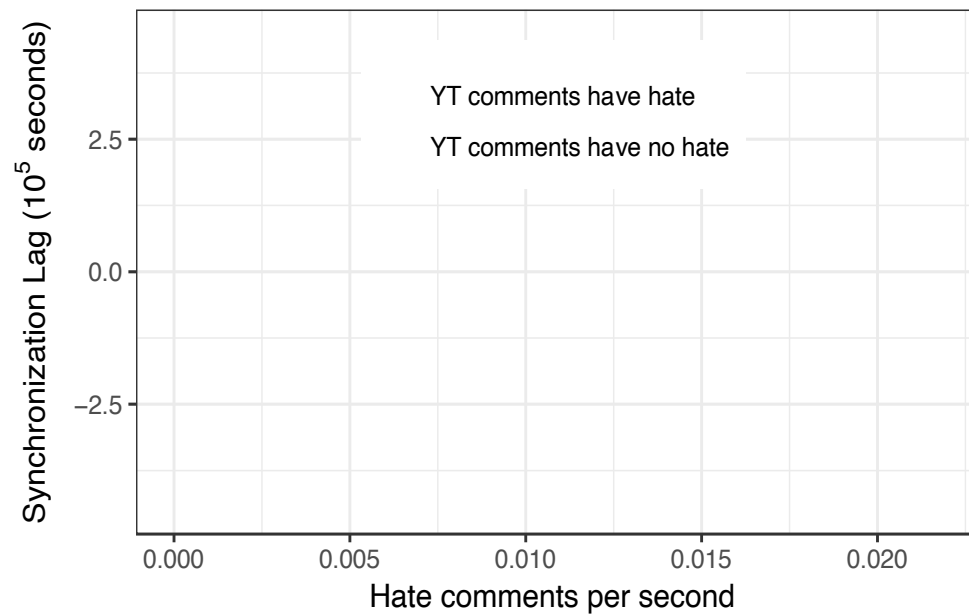


Two series, second randomly shifted from first
by 0.2s on avg



Blue lines → per-sample lag
Red area → density of the lags
Peak of density curve = 0.2s

Validation



Countermeasures?



Countermeasures?

Platforms implement reactive approaches



Countermeasures?

Platforms implement reactive approaches

Mostly rely on feedbacks from the victims of the attacks and manual checks



Countermeasures?

Platforms implement reactive approaches

Mostly rely on feedbacks from the victims of the attacks and manual checks

Moderators going through the flagged content



Countermeasures?

Platforms implement reactive approaches

Mostly rely on feedbacks from the victims of the attacks and manual checks

Moderators going through the flagged content

They rarely use fully automated checks to avoid FP



Countermeasures?

Platforms implement reactive approaches

Mostly rely on feedbacks from the victims of the attacks and manual checks

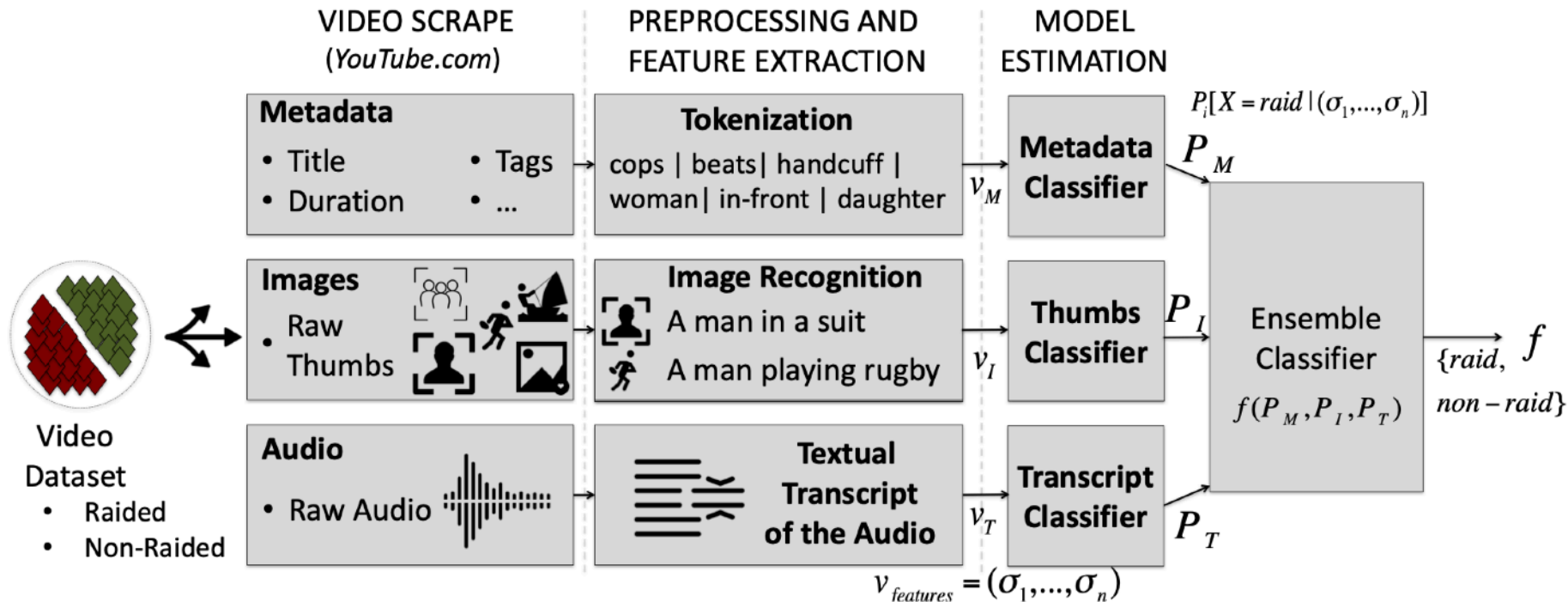
Moderators going through the flagged content

They rarely use fully automated checks to avoid FP

Inherently slow



But... can we predict?



But... can we predict?

Yes, we can!

Ensemble of classifiers
determines the likelihood that
a video will be raided with AUC
up to 94%

VIDEO SCRAPE
(YouTube.com)

PREPROCESSING AND
FEATURE EXTRACTION

MODEL
ESTIMATION

Me

- Title
- Duration

$= \text{raid} \mid (\sigma_1, \dots, \sigma_n)$

P_M

Ensemble
Classifier
 $f(P_M, P_I, P_T)$

$\{ \text{raid}, f_{\text{non-raid}} \}$

Transcript
Classifier

P_T

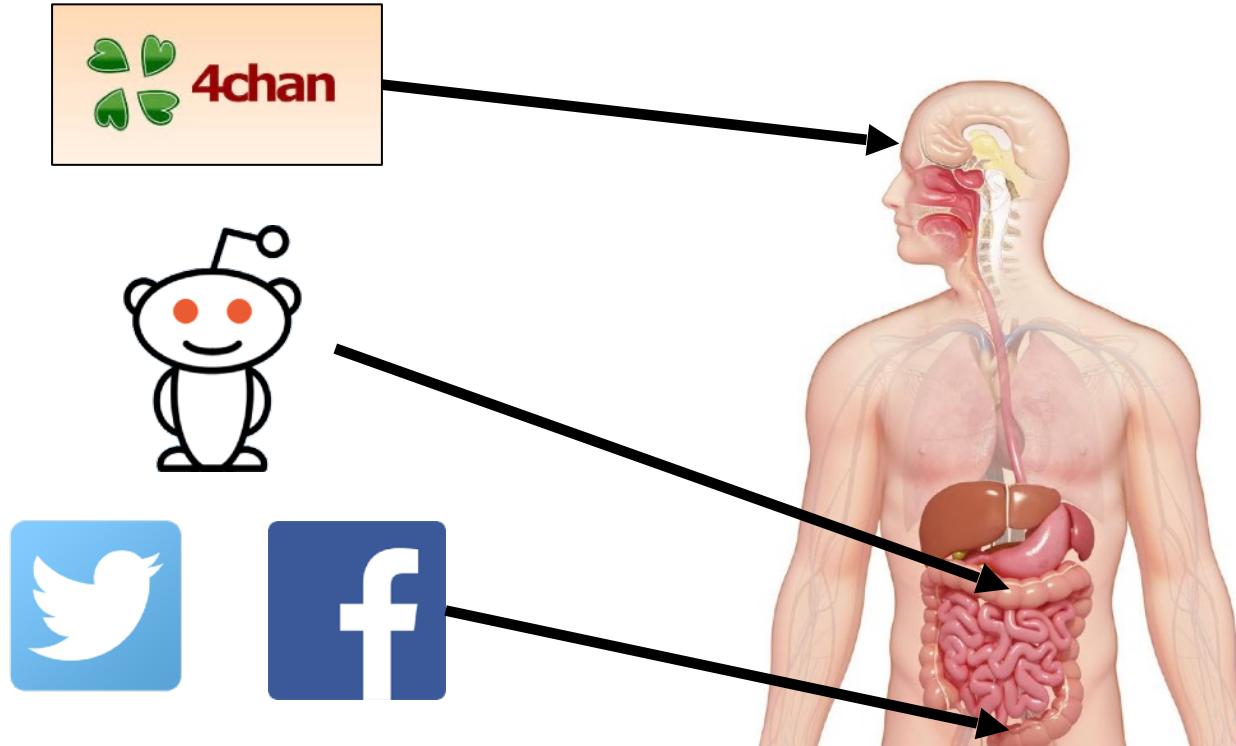
$v_{\text{features}} = (\sigma_1, \dots, \sigma_n)$



Video
Dataset

- Raid
- Non-raid

How do communities influence each other?



The Pizzagate Conspiracy Theory

The saga of 'Pizzagate': The fake story that shows how conspiracy theories spread



By BBC Trending
What's popular and why

🕒 2 December 2016



Share

The Pizzagate Conspiracy Theory

The saga of 'Pizzagate': The fake story that shows how conspiracy theories spread



By BBC Trending
What's popular and why

🕒 2 December 2016



Share



**Data
Provider**

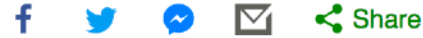
The Pizzagate Conspiracy Theory

The saga of 'Pizzagate': The fake story that shows how conspiracy theories spread



By BBC Trending
What's popular and why

🕒 2 December 2016



Data
Provider



Theory
Generator

The Pizzagate Conspiracy Theory

The saga of 'Pizzagate': The fake story that shows how conspiracy theories spread



By BBC Trending
What's popular and why

🕒 2 December 2016



Data
Provider



Theory
Generator



Theory Incubators & Gateway
to mainstream "world"

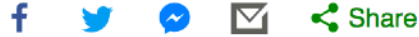
The Pizzagate Conspiracy Theory

The saga of 'Pizzagate': The fake story that shows how conspiracy theories spread



By BBC Trending
What's popular and why

🕒 2 December 2016



Data
Provider



Theory
Generator



Theory Incubators & Gateway
to mainstream "world"



Large-scale
Disseminator

Idea...

Idea...

Study the appearance of alternative and mainstream news URLs within the platforms

Idea...

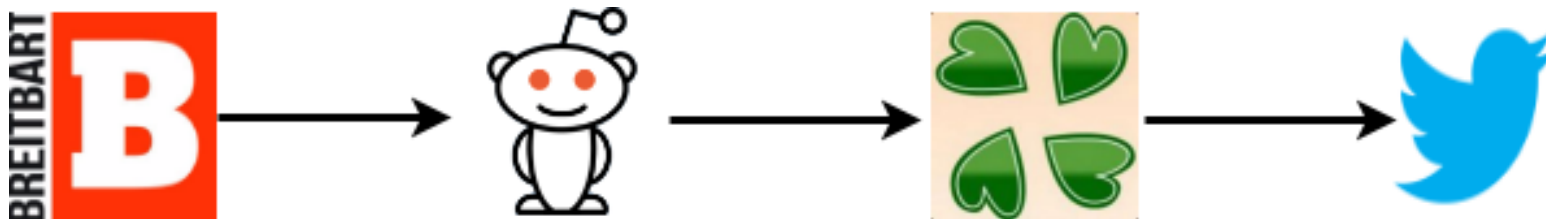
Study the appearance of alternative and mainstream news URLs within the platforms

Build a sequence of appearance for each URL according to the timestamps

Idea...

Study the appearance of alternative and mainstream news URLs within the platforms

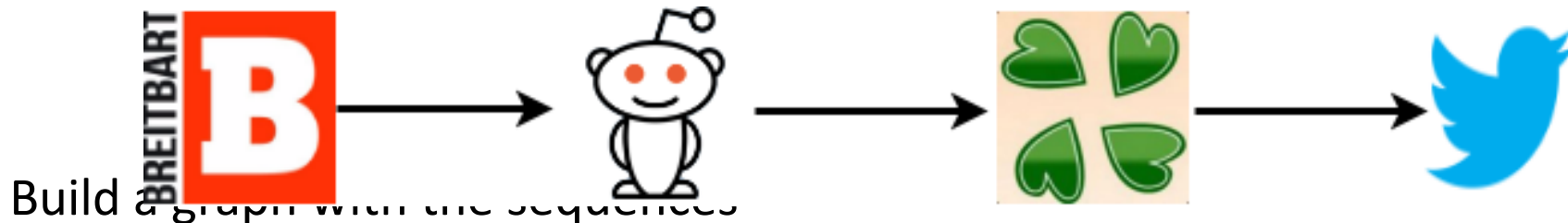
Build a sequence of appearance for each URL according to the timestamps



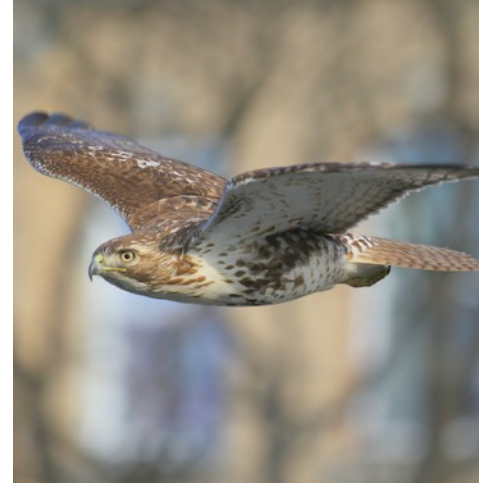
Idea...

Study the appearance of alternative and mainstream news URLs within the platforms

Build a sequence of appearance for each URL according to the timestamps



Quantify Influence? Hawkes Processes!



Quantify Influence? Hawkes Processes!

Assume K processes

Each with a rate of events (i.e., posting of a URL), called the *background rate*



Quantify Influence? Hawkes Processes!

Assume K processes

Each with a rate of events (i.e., posting of a URL), called the *background rate*

An event can cause **impulse responses**

Increases rates of other processes for a period of time



Quantify Influence? Hawkes Processes!

Assume K processes

Each with a rate of events (i.e., posting of a URL), called the *background rate*

An event can cause **impulse responses**

Increases rates of other processes for a period of time

Confidence about the number of events caused by another event on the source process (weight)

Reveals causal relationships



For Our Purposes...

Hawkes model with 8 processes

- One for each platform/community

- Distinct model for each URL

Fit each model with Gibbs sampling

Calculate the percentage of events **created because** of events happened in each of the other processes

How Communities Influence Each Other?

Twitter top influencers for alternative URLs

- /r/The_Donald (2.72%)
- /pol/ (1.96%)
- /r/politics (1.1%)

Twitter top influencers for mainstream URLs

- /r/politics (4.29%)
- /pol/ (3.01%)
- /r/The_Donald (2.97%)

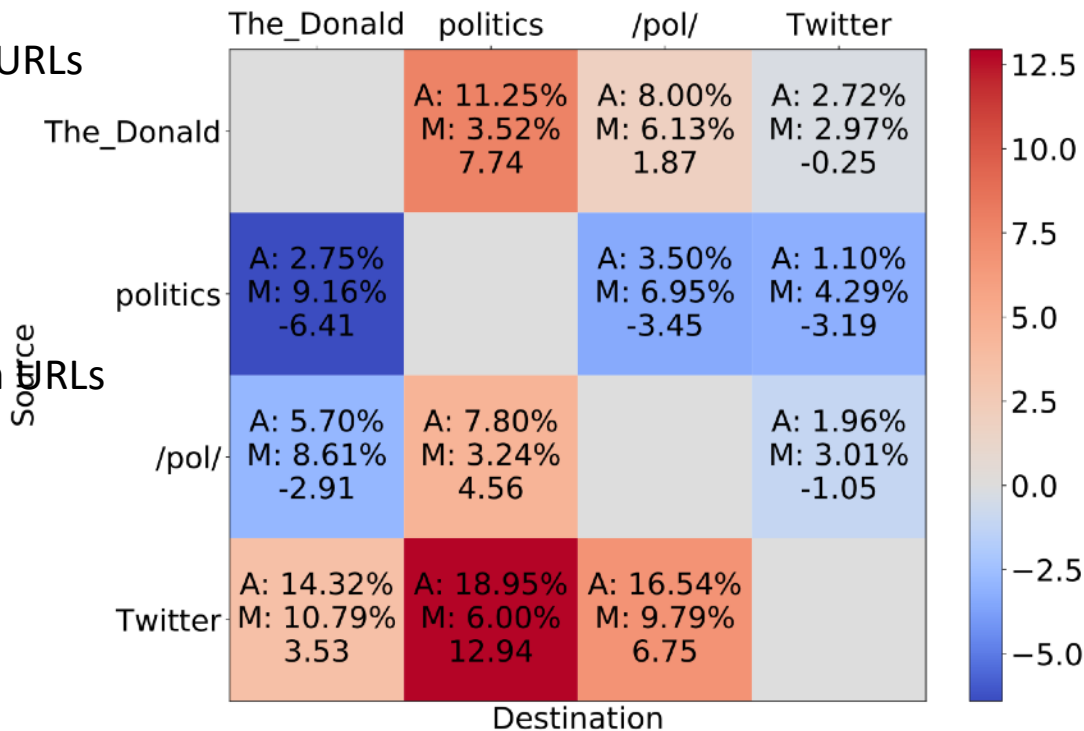
How Communities Influence Each Other?

Twitter top influencers for alternative URLs

- /r/The_Donald (2.72%)
- /pol/ (1.96%)
- /r/politics (1.1%)

Twitter top influencers for mainstream URLs

- /r/politics (4.29%)
- /pol/ (3.01%)
- /r/The_Donald (2.97%)



How Communities Influence Each Other?

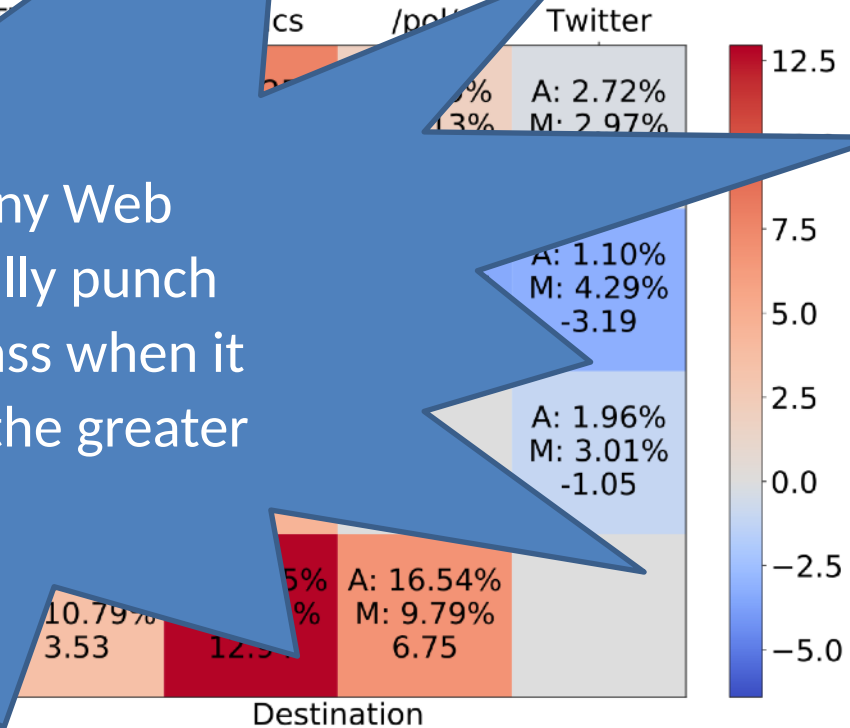
Twitter top influencers alternative

- /r/The_Donald
- /pol/ (1.96%)

Twitter top influencers

- /r/The_Donald

These seemingly tiny Web communities can really punch above their weight class when it comes to influencing the greater Web







Memes are fun!

Not always though...



Not always though...



Not always though...



Not always though...



Memes in politics

Sport

Culture

Lifestyle

More ▾

The Guardian

US midterms 2018 [Business](#) Tech Science

Annotated: the Trump memes stuck to Cesar Sayoc's van

The vehicle seized by law enforcement in Florida was covered in strange propaganda



▲ Cesar Sayoc's van is seen in Boca Raton, Florida, on 18 October 2018 in this picture obtained from social media. Photograph: Social Media/Ed Kennedy via Reuters

most viewed in US



Pittsburgh synagogue shooting: suspect held after at least four killed



'I bet \$500 they are lying': Trump fans sceptical about pipe bomb arrest



'Why is that racist?' Trump greets young black leaders with freewheeling rant



Barack Obama takes aim at Donald Trump for 'making stuff up'



Muncy hits walk-off homer as Dodgers win longest ever World Series game



Sport

Culture

Lifestyle

More ▾

The Guardian

US midterms 2018 [Business](#) Tech Science

Annotated Trump memo to Cesar Sayoc

Memes have become a popular, and seemingly effective, method to transmit ideology.

Memes have been weaponized

most vis

gh synagogue

aying':
sceptical about
st

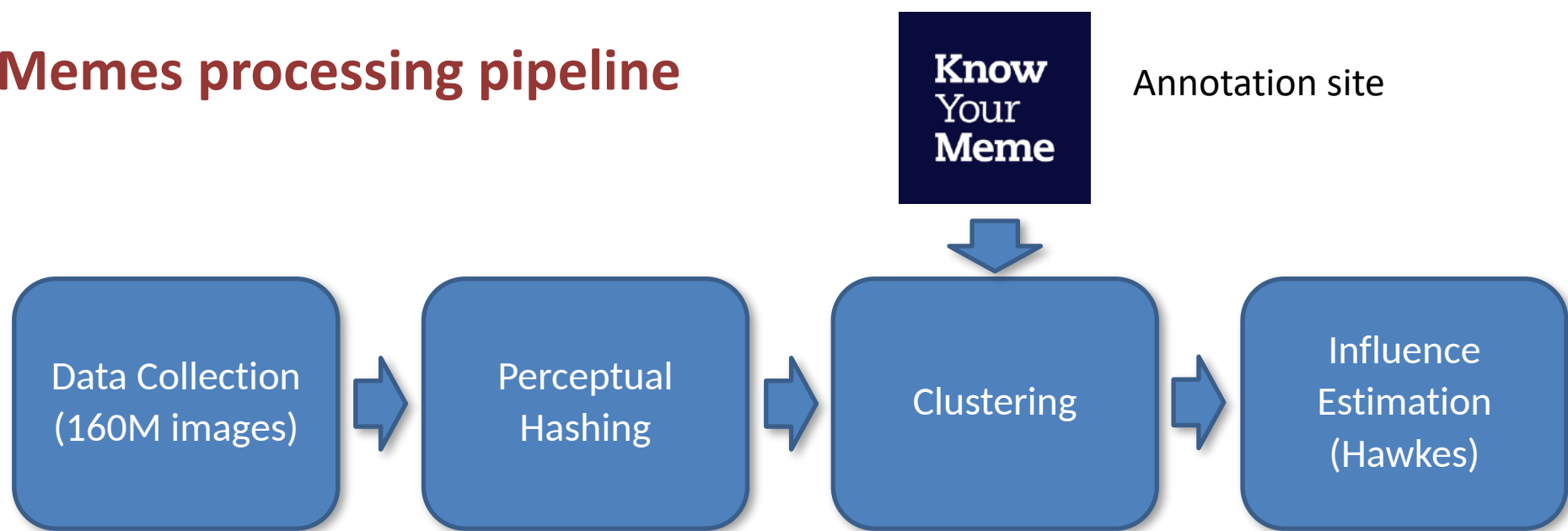
ist: Trump
young black leaders
wheeling rant

s aim at
Donald Trump
stuff up'

Muncy hits walk-off homer as
Dodgers win longest ever
World Series game

▲ Cesar Sayoc's van is seen in Boca Raton, Florida, on 18 October 2018 in this picture obtained from social media.
Photograph: Social Media/Ed Kennedy via Reuters

Memes processing pipeline



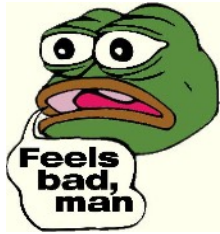
Code for our analysis pipeline (and data) available at
https://github.com/memespaper/memes_pipeline

Examples of clusters

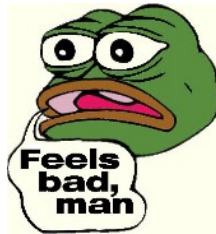


Top memes per social network

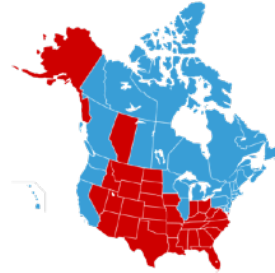
4chan (/pol/)



Reddit



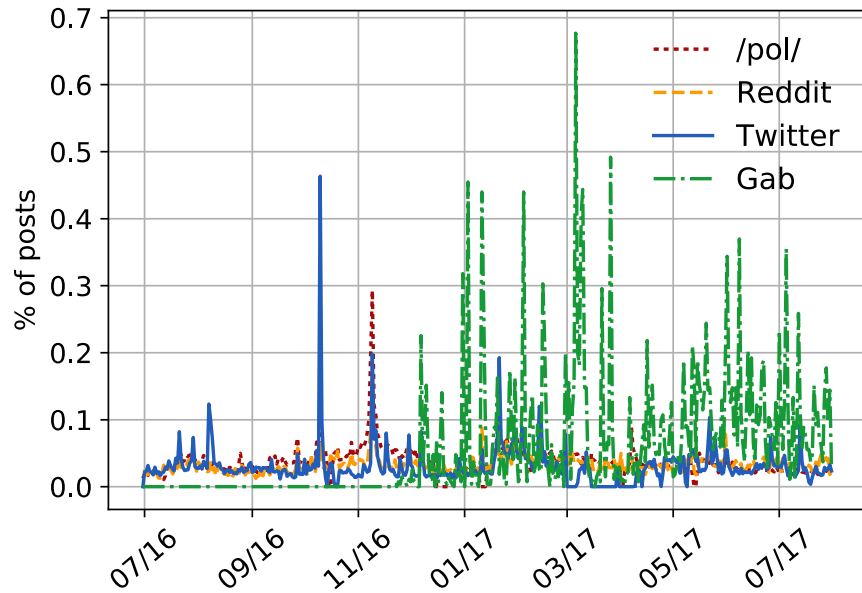
Gab



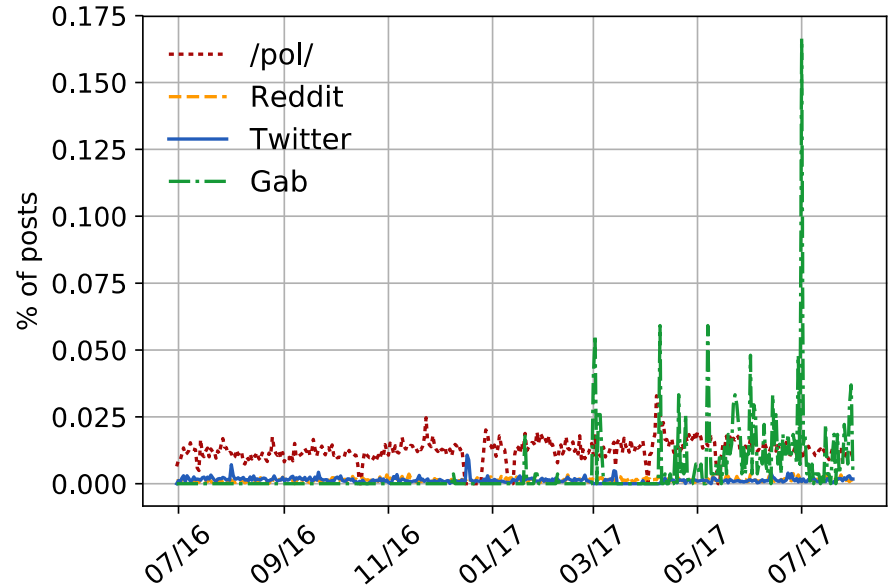
Twitter



How are memes shared over time?



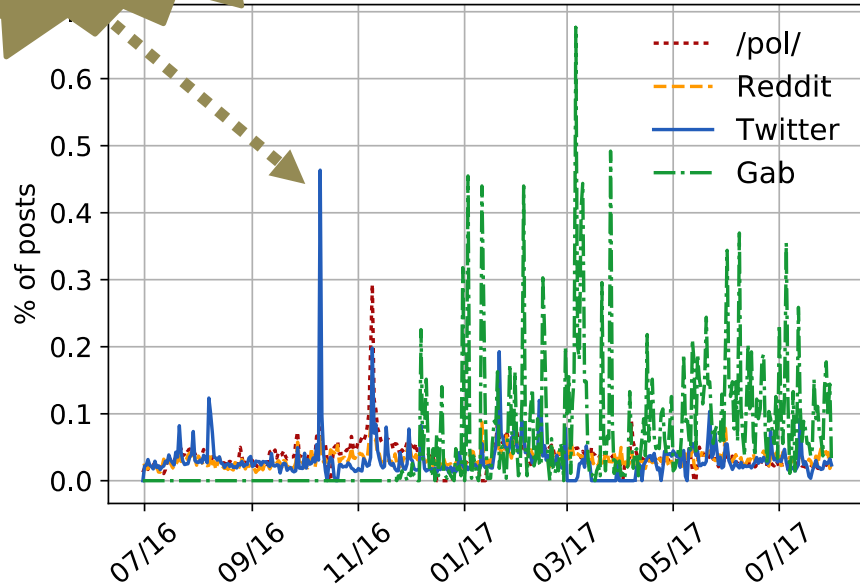
Political Memes



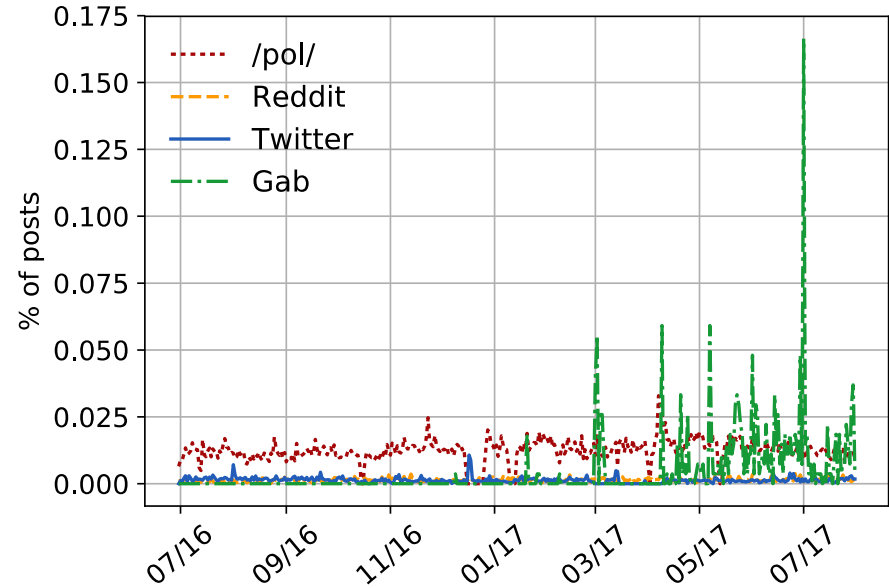
Racist Memes

How are memes shared over time?

2nd US
presidential
debate

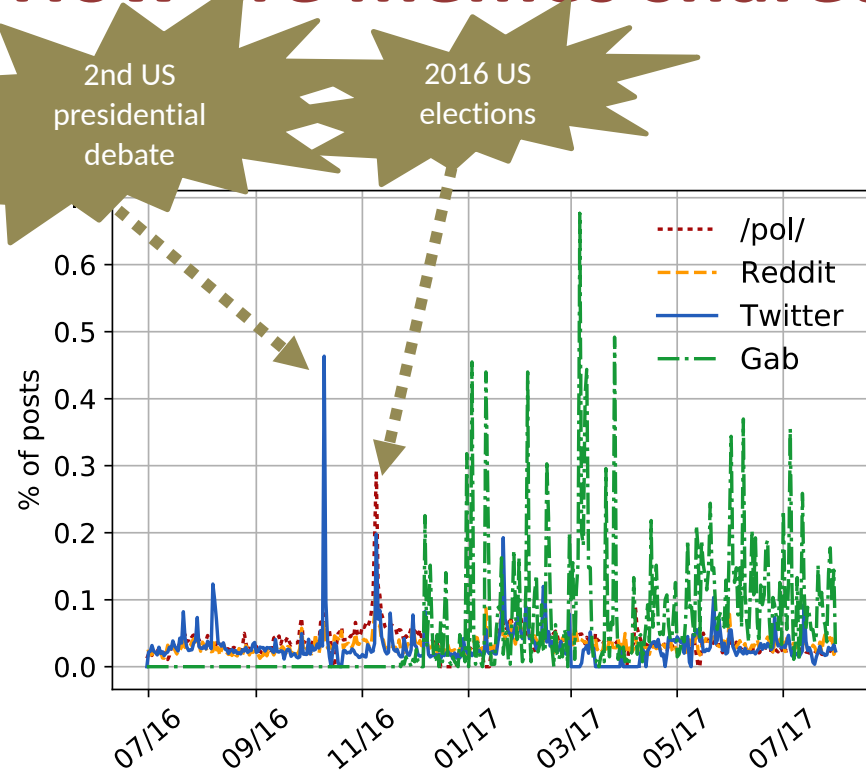


Political Memes

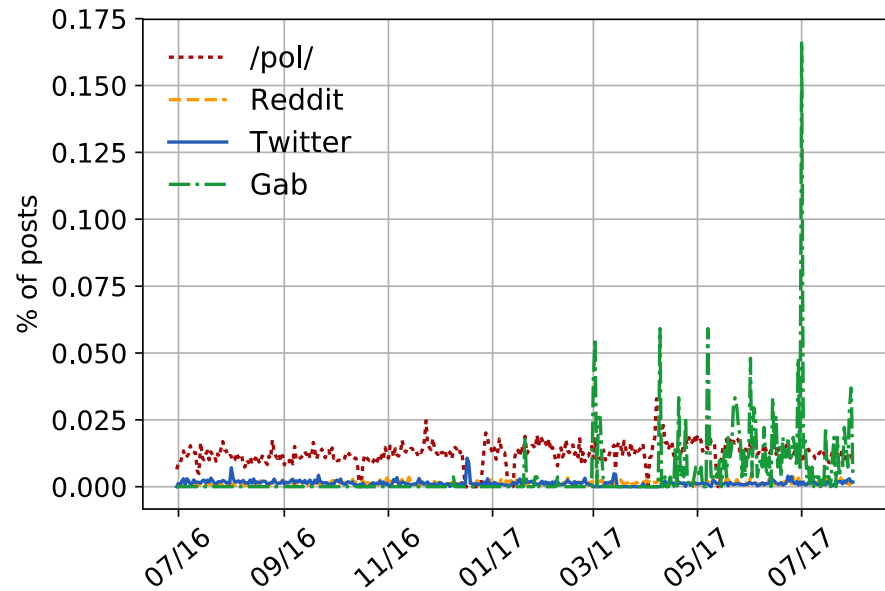


Racist Memes

How are memes shared over time?

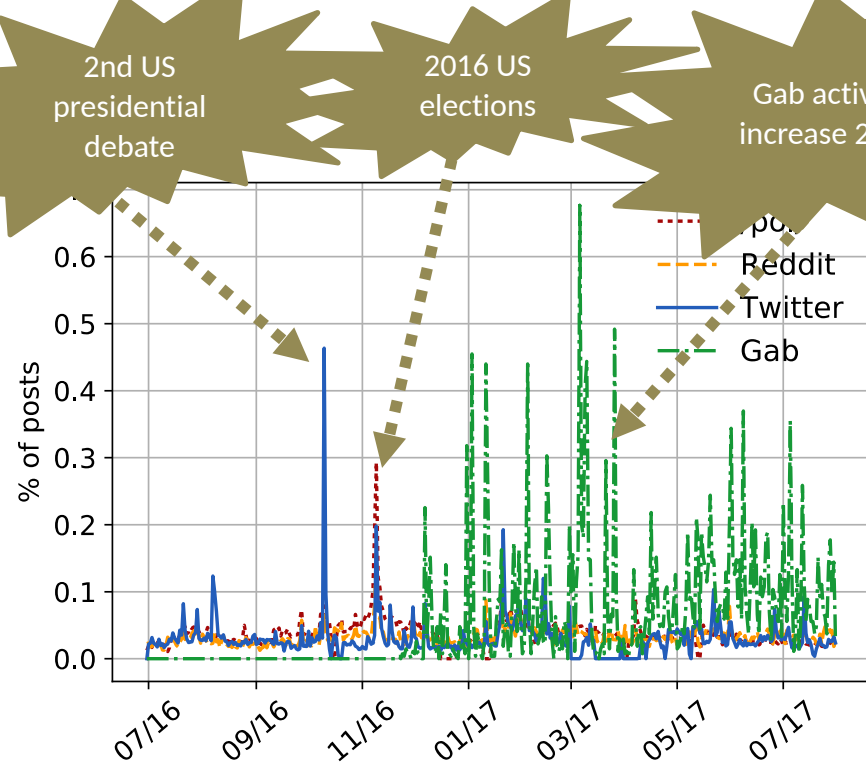


Political Memes

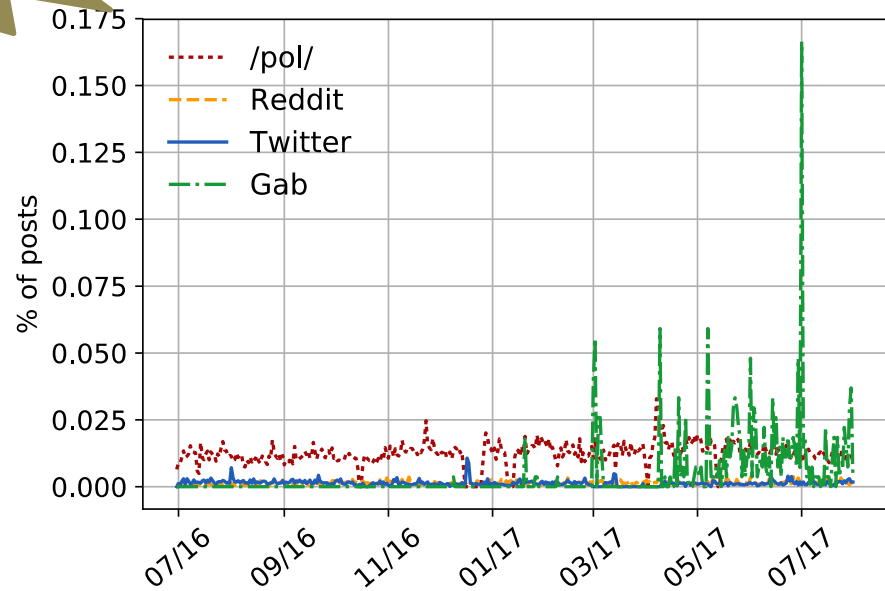


Racist Memes

How are memes shared over time?

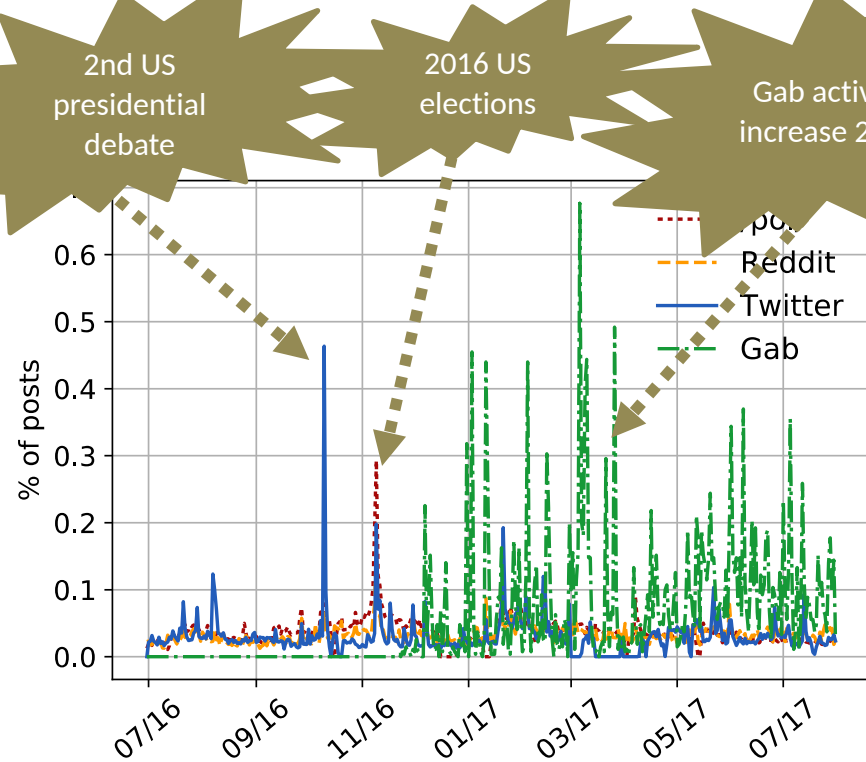


Political Memes

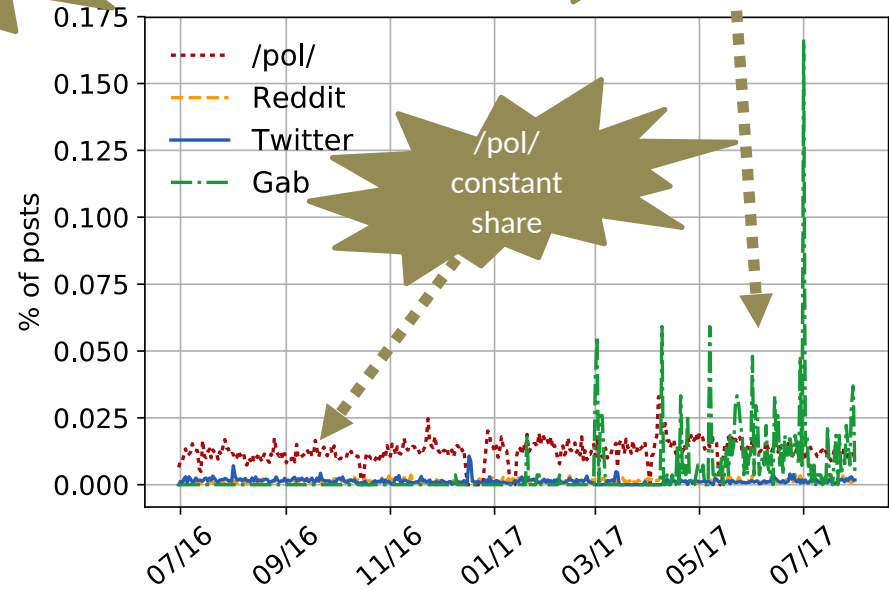


Racist Memes

How are memes shared over time?



Political Memes



Racist Memes

Communities' influence (racist memes)

/pol/ is most influential in terms of spreading racist memes

		/pol/	Reddit	Twitter	Gab	T_D
Source	/pol/	R: 99.44% NR: 97.12%*	R: 6.89% NR: 3.95%*	R: 3.94% NR: 2.93%*	R: 18.93% NR: 12.91%	R: 14.17% NR: 16.41%*
	Reddit	R: 0.29% NR: 1.28%*	R: 88.95% NR: 90.87%*	R: 2.79% NR: 4.64%*	R: 1.25% NR: 9.42%	R: 9.84% NR: 9.02%*
	Twitter	R: 0.15% NR: 0.79%	R: 1.75% NR: 2.85%*	R: 92.95% NR: 90.96%*	R: 0.66% NR: 8.16%	R: 1.37% NR: 3.66%*
	Gab	R: 0.04% NR: 0.09%	R: 0.65% NR: 0.16%	R: 0.10% NR: 0.20%	R: 76.67% NR: 59.72%	R: 0.08% NR: 0.58%
	T_D	R: 0.07% NR: 0.73%*	R: 1.75% NR: 2.18%*	R: 0.23% NR: 1.27%*	R: 2.49% NR: 9.79%	R: 74.53% NR: 70.32%*
		Destination				

Communities' efficiency (racist memes)

If we look at the influence normalized to the number of memes posted, the The_Donald is most efficient in terms of disseminating memes

		/pol/	Reddit	Twitter	Gab	T_D	Total Ext
Source	/pol/	R: 99.4 NR: 97.1*	R: 0.5 NR: 1.5*	R: 0.3 NR: 1.4*	R: 0.2 NR: 0.4	R: 0.1 NR: 0.9*	R: 1.1 NR: 4.2
	Reddit	R: 4.2 NR: 3.3*	R: 89.0 NR: 90.9*	R: 2.7 NR: 5.7*	R: 0.2 NR: 0.7	R: 1.4 NR: 1.3*	R: 8.5 NR: 11.1
	Twitter	R: 2.2 NR: 1.7	R: 1.8 NR: 2.3*	R: 92.9 NR: 91.0*	R: 0.1 NR: 0.5	R: 0.2 NR: 0.4*	R: 4.4 NR: 4.9
	Gab	R: 4.5 NR: 3.0	R: 4.9 NR: 2.0	R: 0.7 NR: 3.2	R: 76.7 NR: 59.7	R: 0.1 NR: 1.1	R: 10.3 NR: 9.3
	T_D	R: 7.5 NR: 13.6*	R: 12.1 NR: 15.5*	R: 1.5 NR: 11.1*	R: 2.3 NR: 5.3	R: 74.5 NR: 70.3*	R: 23.4 NR: 45.6
		Destination					

What is a meme anyway?

It should meet two conditions:

1. Shared by more than one user on social media
2. Present at least one variation on social media





What visual elements make a meme go viral?

1. Composition
2. Subjects
3. Target audience



i) A viral meme.



ii) A non-viral meme

Twitter,
Reddit, 4chan,
and Gab

Collecting memes

Code Book

Labeling $O(100)$ memes

Classifiers

Predicting virality

- Composition and subject matters. A close-up scale, contain characters, and include positive or negative emotions may go viral
- Random Forest achieves 0.866 Area Under the Curve (AUC), which can distinguish viral and non-viral memes from both fringe and mainstream platforms, such as Twitter and Reddit
- See: <https://arxiv.org/pdf/2101.06535.pdf>

What about state-level adversaries?

What about state-level adversaries?

Raise your hand if you're under investigation!



a)



b)



c)

Research Questions

- How do state-sponsored actors operate and evolve?
- How does the behavior of state-sponsored trolls compare to random users?
- More importantly, what was their influence on the Web with respect to the dissemination of news?

What are they posing as?

News
accounts?

Nudging
users to
follow them

Description			
Word	(%)	Word bigram	(%)
news	10.7%	follow me	7.8%
follow	10.7%	breaking news	2.6%
conservative	8.1%	news aus	2.1%
trump	7.8%	uns in	2.1%
und	6.2%	deiner stdt	2.1%
maga	5.9%	die news	2.1%
love	5.8%	wichtige und	2.1%
us	5.3%	nachrichten aus	2.1%
die	5.0%	aus deiner	2.1%
nachrichten	4.3%	die dn	2.1%

Trump
supporters

Where are they allegedly posting from?



What hashtags they shared?

Trolls			Baseline		
Hashtag	(%) Hashtag	(%)	Hashtag	(%) Hashtag	(%)
news	7.2% US	0.7%	iHeartAwards	1.8% UrbanAttires	0.6%
politics	2.6% tcot	0.6%	BestFanArmy	1.6% Vacature	0.6%
sports	2.1% PJNET	0.6%	Harmonizers	1.0% mPlusPlaces	0.6%
business	1.4% entertainment	0.5%	iOSApp	0.9% job	0.5%
money	1.3% top	0.5%	JouwBaan	0.9% Directioners	0.5%
world	1.2% topNews	0.5%	vacature	0.9% JIMIN	0.5%
MAGA	0.8% ISIS	0.4%	KCA	0.9% PRODUCE101	0.5%
health	0.8% Merkelmussbleiben	0.4%	Psychic	0.8% VoteMainFPP	0.5%
local	0.7% IslamKills	0.4%	RT	0.8% Werk	0.4%
BlackLivesMatter	0.7% breaking	0.4%	Libertad2016	0.6% dts	0.4%

Do Russian trolls change their screen names?

9% of the Russian troll accounts changed their screen name

Up to 4 times per account

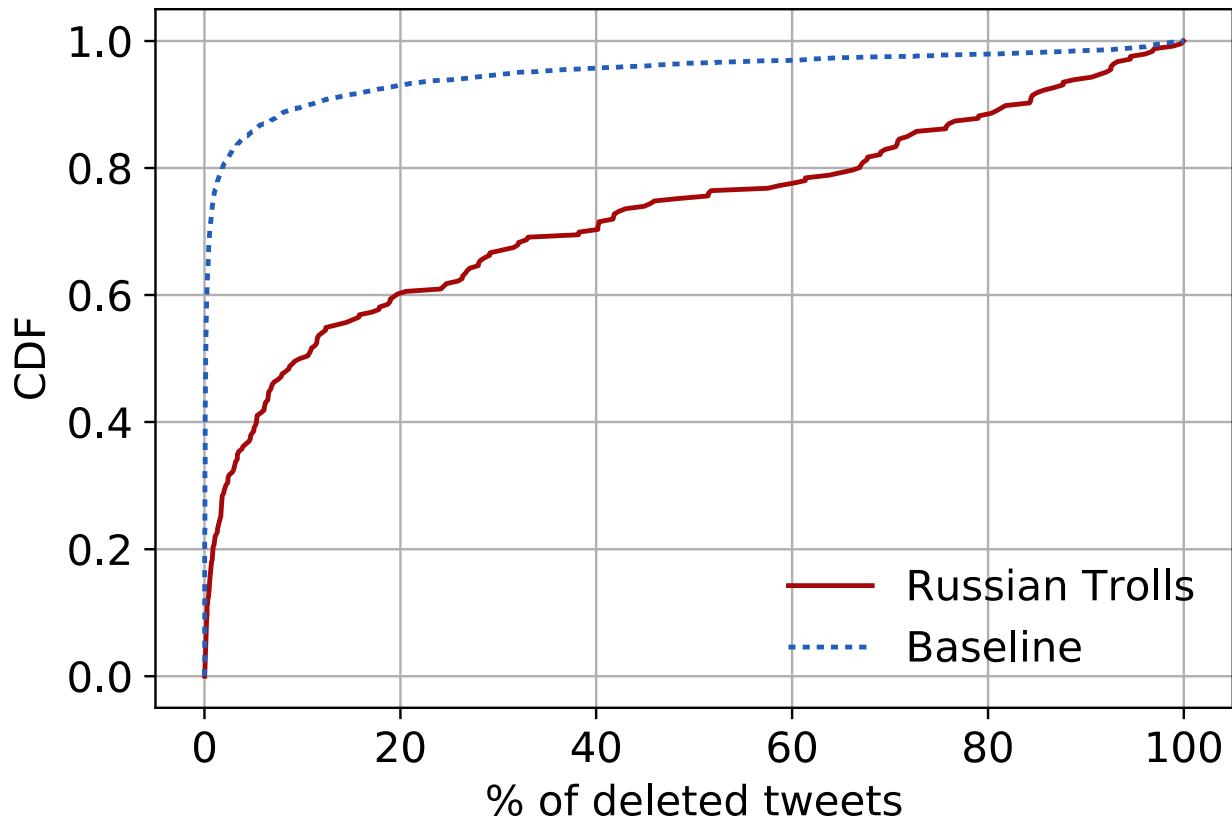
E.g., from “OnlineHouston” to “HoustonTopNews”

Clear attempt to pose as local news outlet

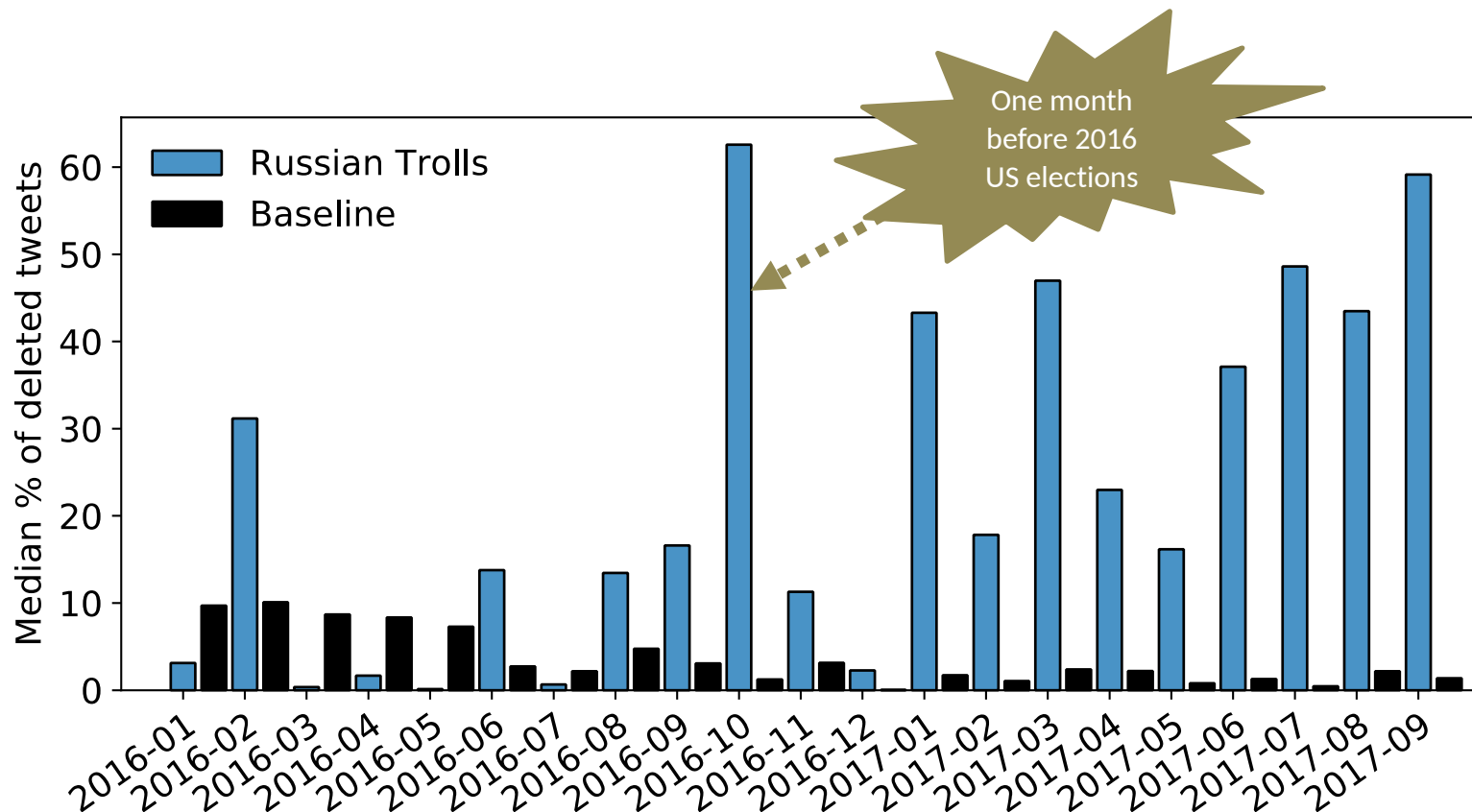
In our baseline dataset 19% of the accounts changed their screen name

Up to 11 times per account

Do Russian trolls delete their tweets?



When did Russian trolls deleted their tweets?



Influence

1K trolls caused 2.6% of all Russian state-sponsored news outlets URLs (i.e., RT) on Twitter's 1%.

1K trolls caused 0.6% of all other news outlets URLs on Twitter's 1%.

	/pol/	Reddit	Twitter	Trolls
/pol/		R: 5.74% O: 8.15% -2.41	R: 0.71% O: 1.17% -0.46	R: 5.32% O: 9.07% -3.75
Reddit	R: 4.78% O: 46.78% -41.99		R: 5.57% O: 12.22% -6.64	R: 13.20% O: 57.15% -43.95
Twitter	R: 24.90% O: 9.14% 15.75	R: 16.66% O: 10.49% 6.17		R: 43.84% O: 51.53% -7.68
Trolls	R: 1.38% O: 0.72% 0.66	R: 3.13% O: 0.62% 2.51	R: 2.69% O: 0.61% 2.07	
Destination				

Main Findings

Differences in the use of the Twitter platform between trolls and random users

Trolls seem to reset their “personas” by changing names and deleting tweets

Particularly influential in spreading Russian state-sponsored URLs on Twitter and other platforms

Manosphere

/ˈmɑ-nə-sfɪə/

Web-based anti-feminist movements roughly focused on "men's issues," which have seen significant growth over the past years

INCEL



The “black-pill”

“... ugly, 'genetically inferior' men have no chance of getting laid (...) as women are assumed to choose based on looks rather than personality or effort.”

Ideas espoused in the community are tightly linked to real world acts of misogynist terrorism

2014 Isla Vista killings



Isla Vista (southern California)

Location	Isla Vista, California, United States
Coordinates	 34.412°N 119.859°W
Date	May 23, 2014 ~9:27 p.m. (UTC-8:00)
Target	Students of the University of California, Santa Barbara
Attack type	Spree killing · murder-suicide · drive-by shooting · stabbing · mass shooting · vehicle-ramming attack
Weapons	Two knives · Glock 34 handgun · Two SIG Sauer P226 handguns · BMW 328i Coupé
Deaths	7 (3 by stabbing, 4 by gunfire including the perpetrator)
Injured	14 (7 by gunfire, 7 struck by motor vehicle)
Perpetrator	Elliot Rodger
Motive	Revenge for perceived sexual and social rejection , misogyny

Idea

Provide a birds-eye view of the evolution of the Manosphere

Data

52 subreddits

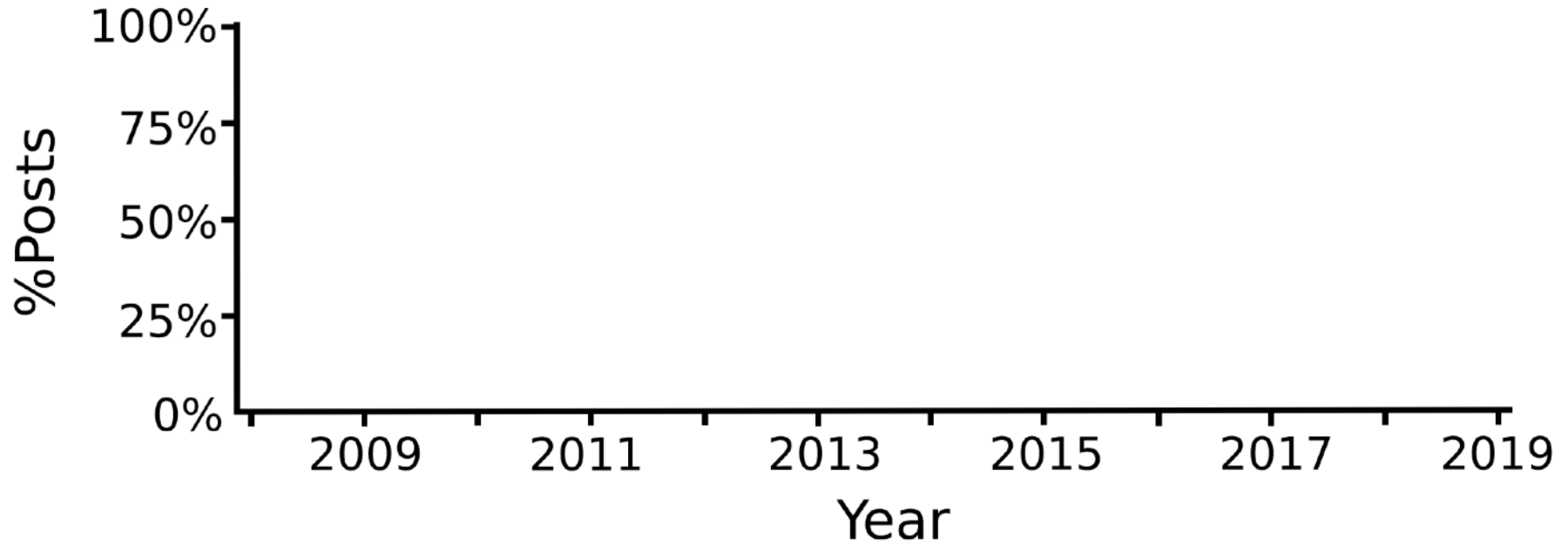
6 standalone forums

835k users

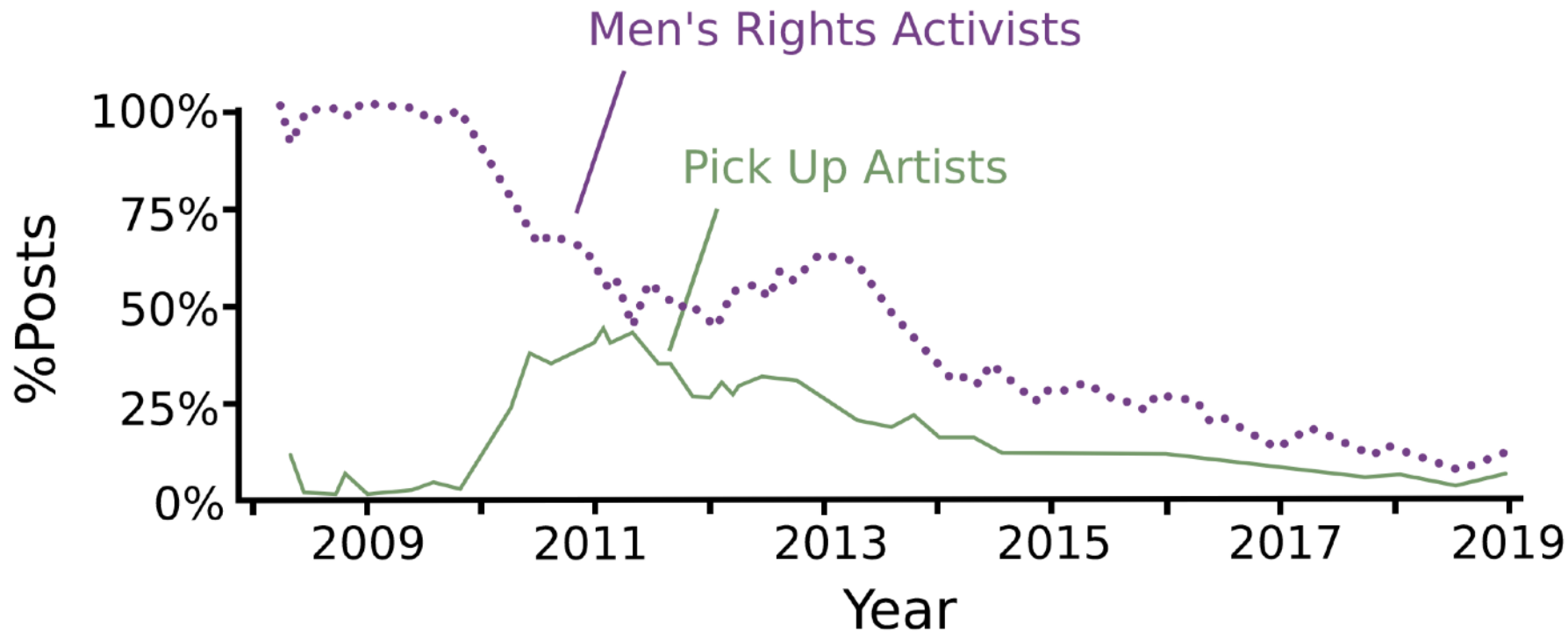
22M posts

RQ1 How has the popularity and the levels of activity of the different Manosphere communities evolved?

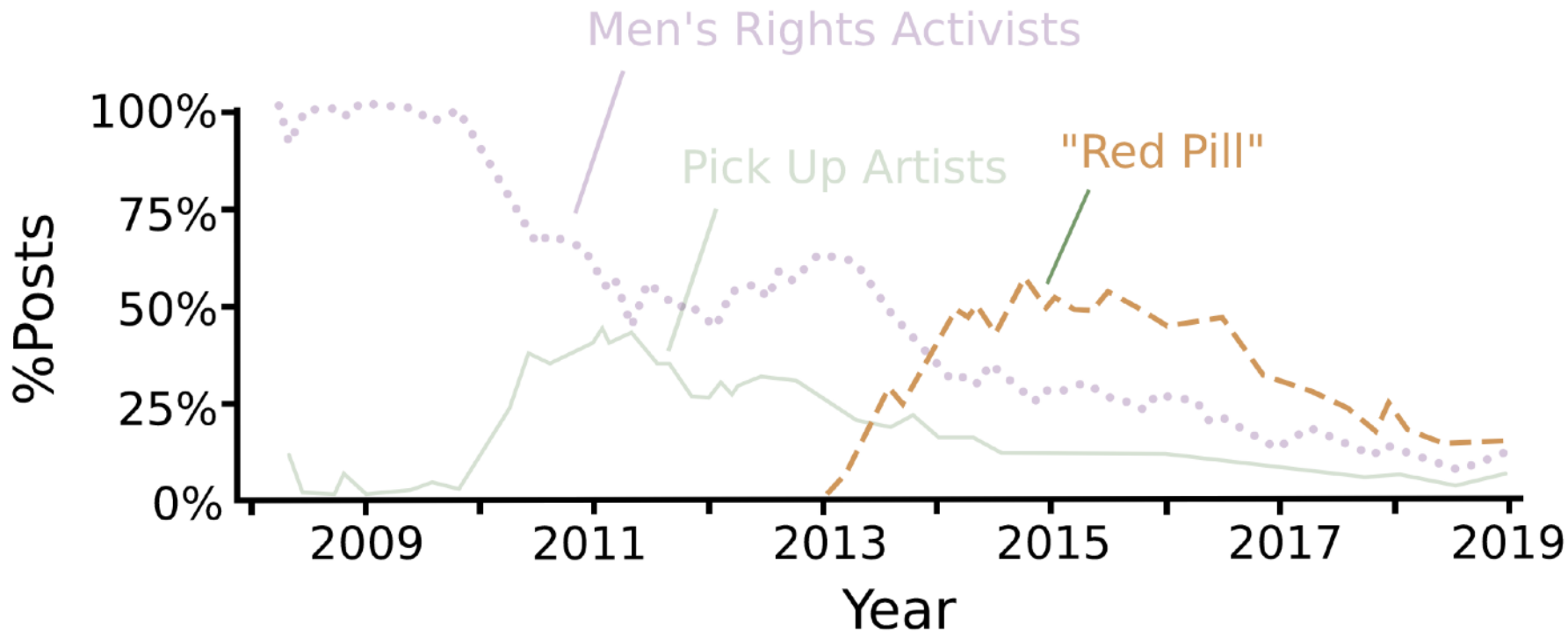
Activity on Reddit



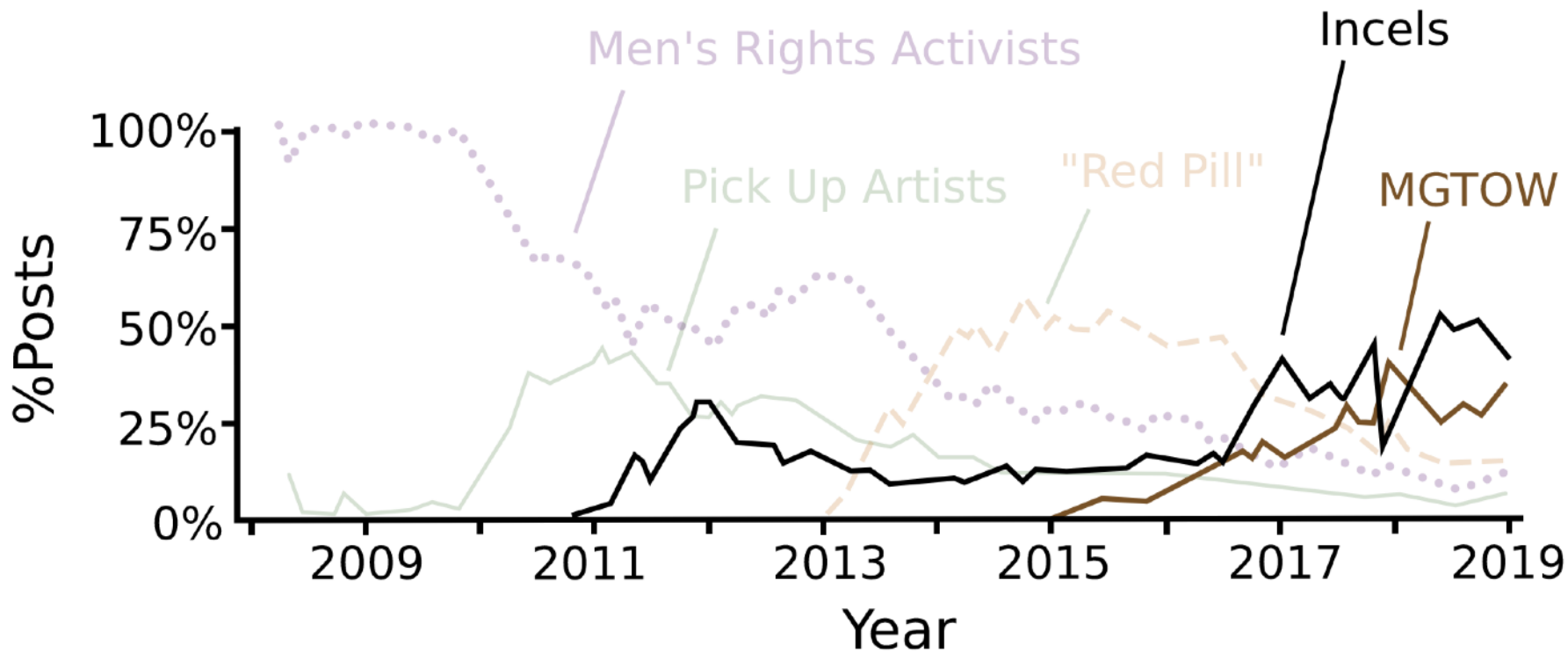
Activity on Reddit



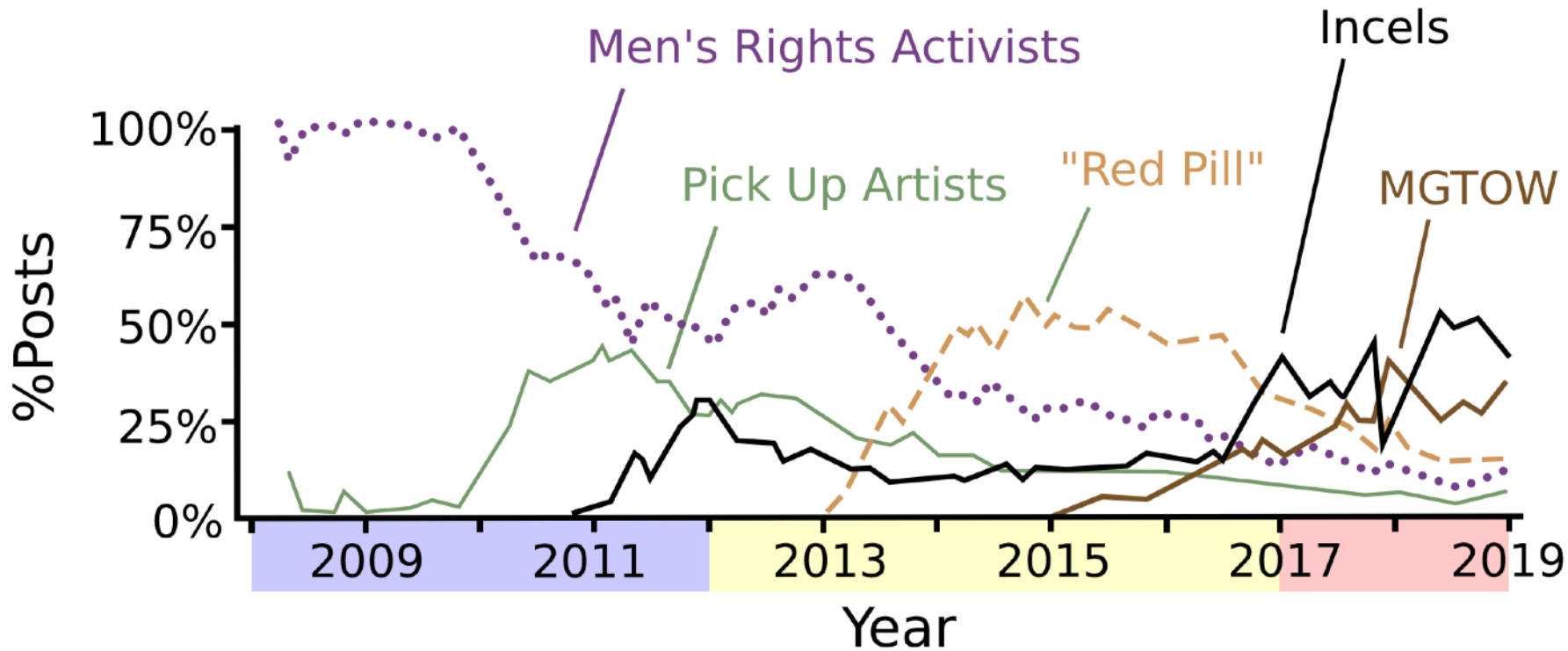
Activity on Reddit



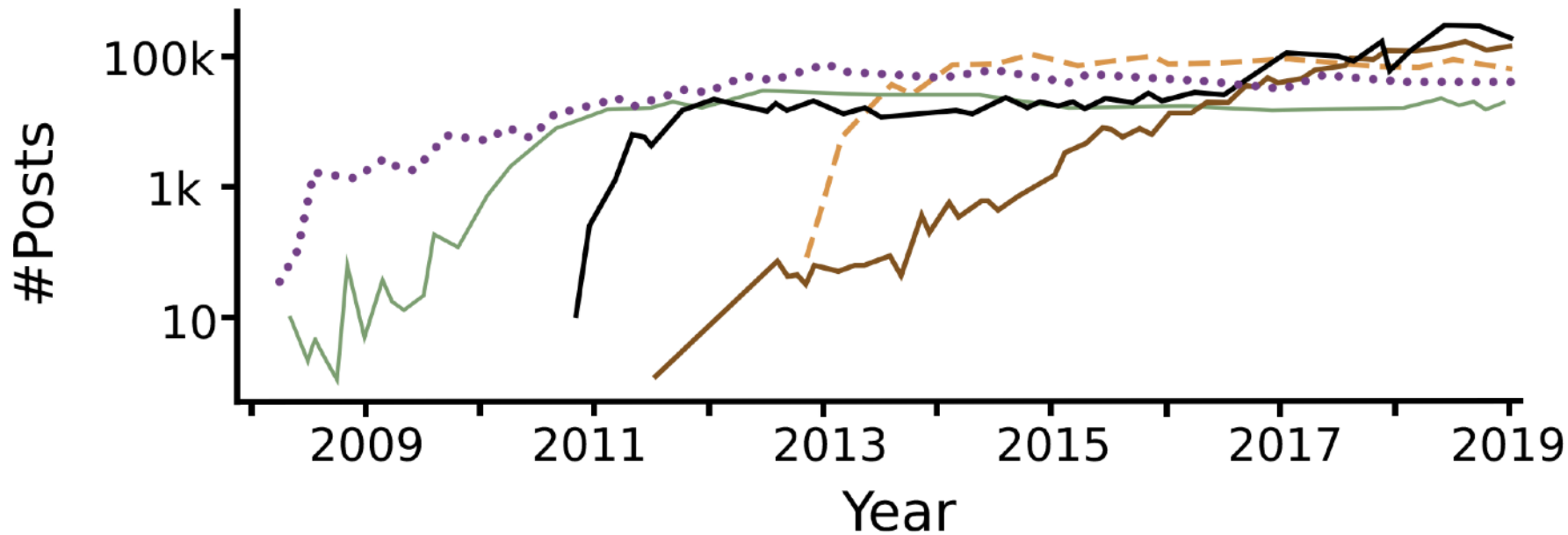
Activity on Reddit



Activity on Reddit



Activity on Reddit

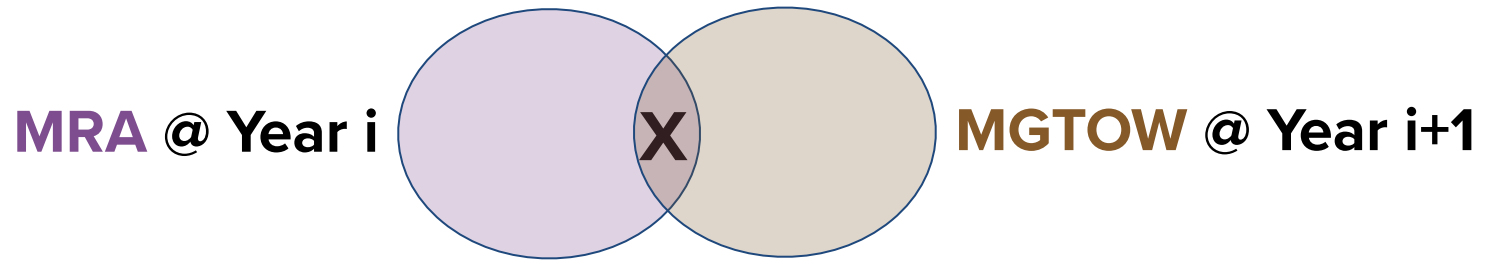


Takeaway #1

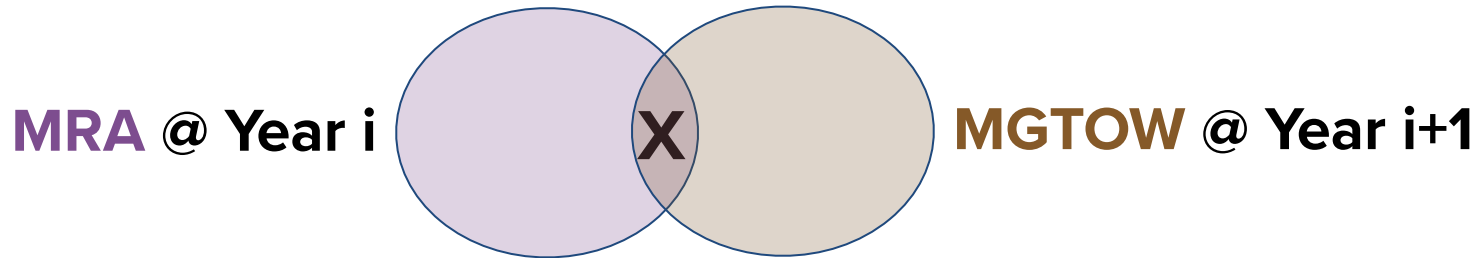
The Manosphere is dynamic! Older communities (MRAs and PUAs) are being overshadowed by newer ones (Incels and MGTOWs).

RQ2

Has there been substantial migration of, or intersection in, users across communities?

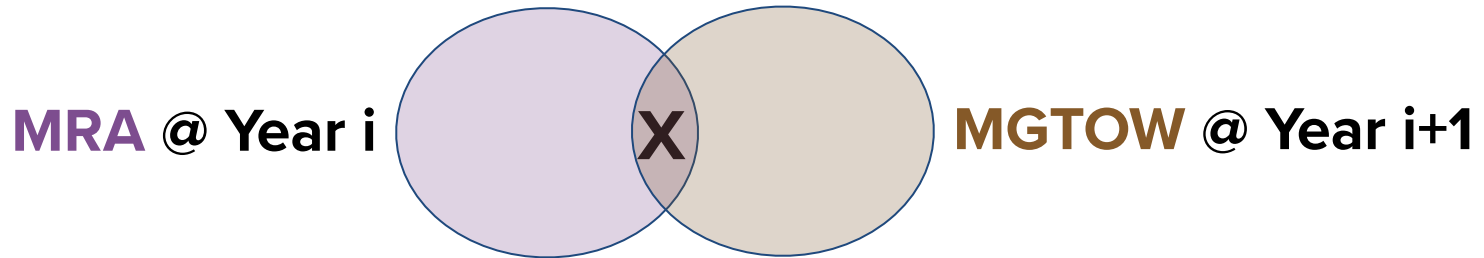


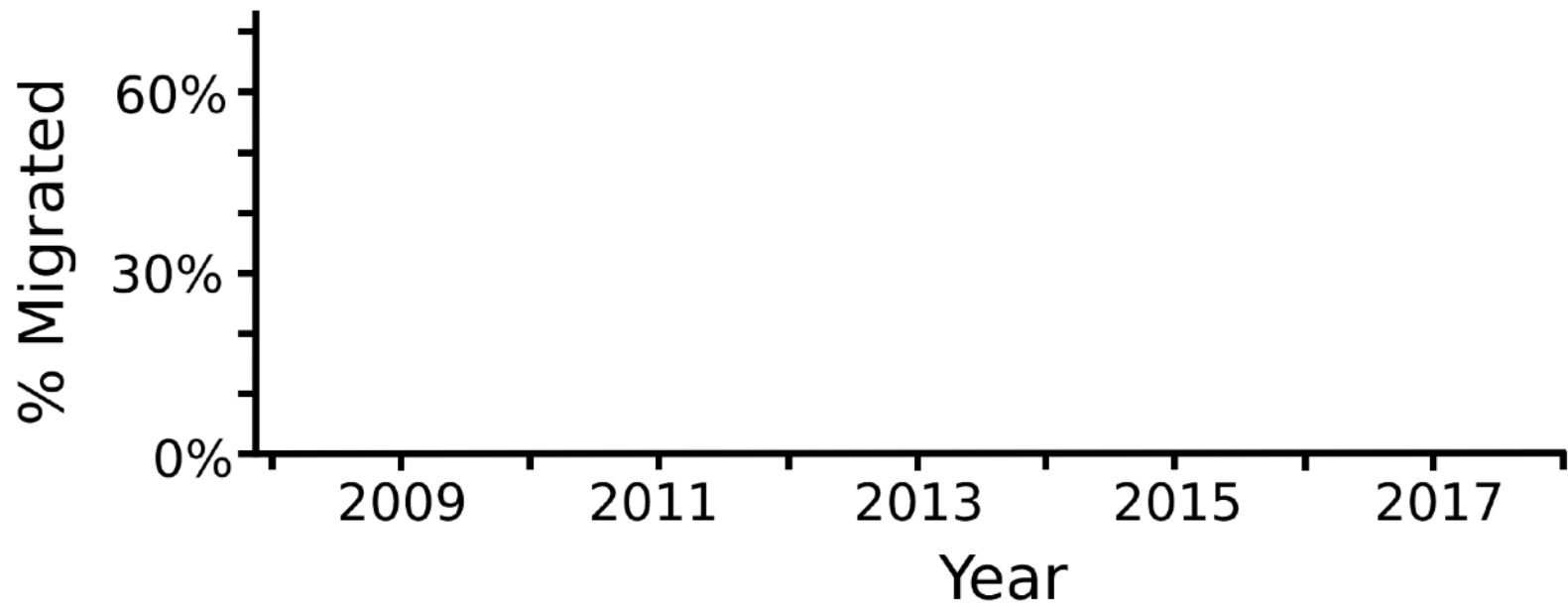
$$\text{Migration Relative to Source} = \frac{X}{\text{MRA @ Year } i}$$

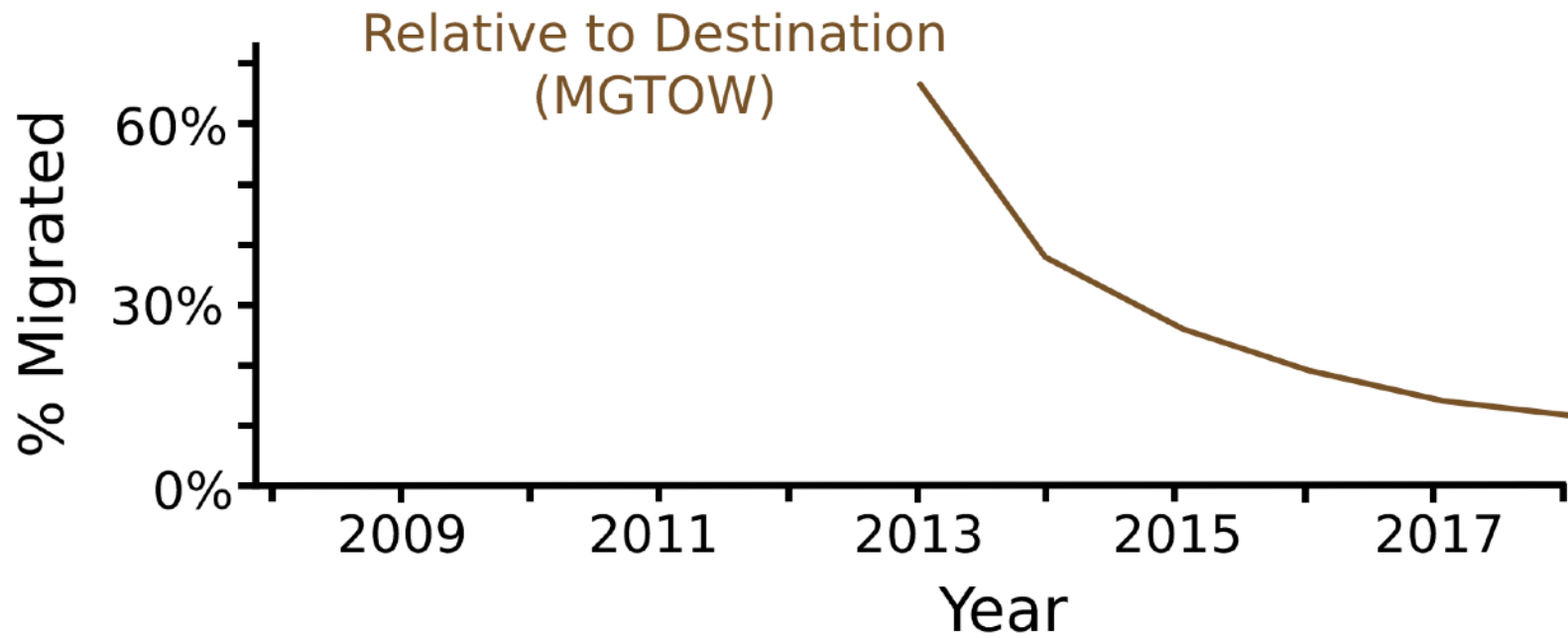


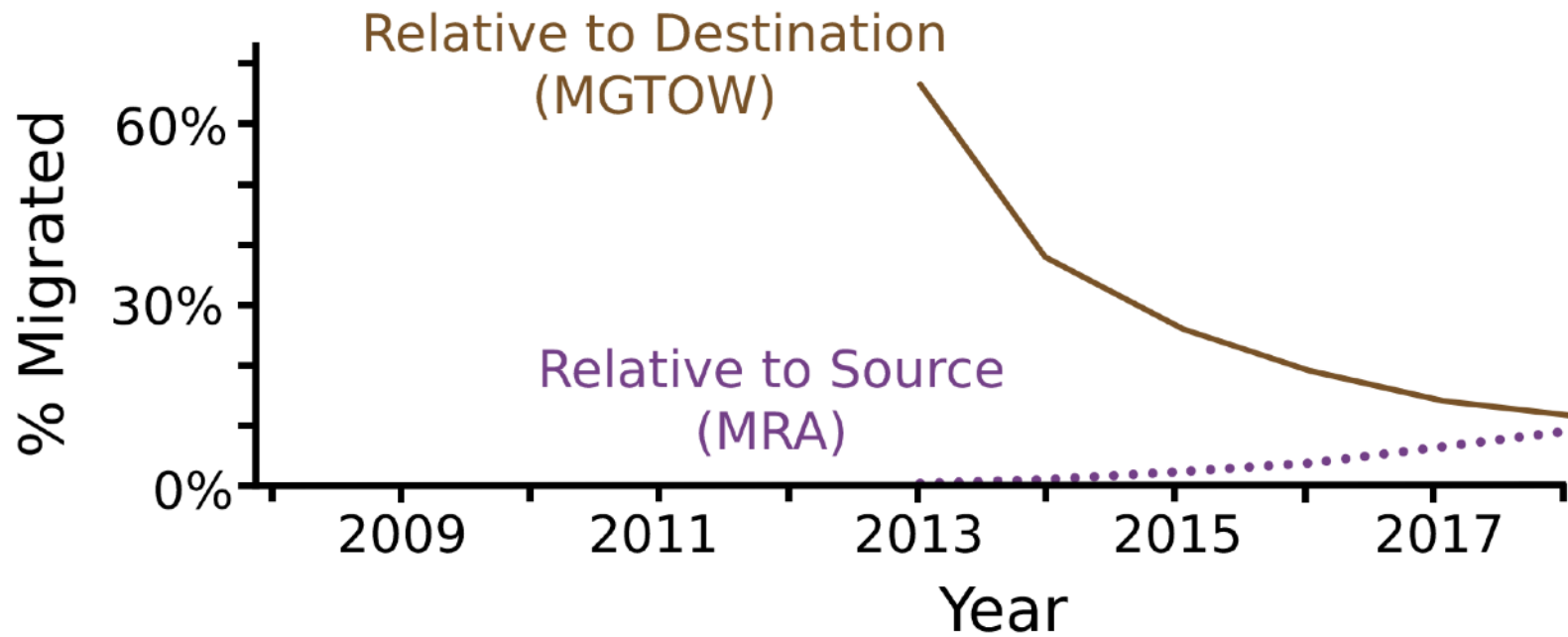
$$\text{Migration Relative to Source} = \frac{X}{\text{MRA @ Year } i}$$

$$\text{Migration Relative to Destination} = \frac{X}{\text{MGTO}W @ \text{Year } i + 1}$$









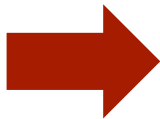
Takeaway #2

Communities in the Manosphere are connected by members who participate in multiple communities and migrate between them.

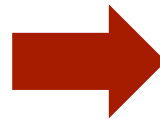
RQ3

Has speech become more toxic and/or misogynistic over time?

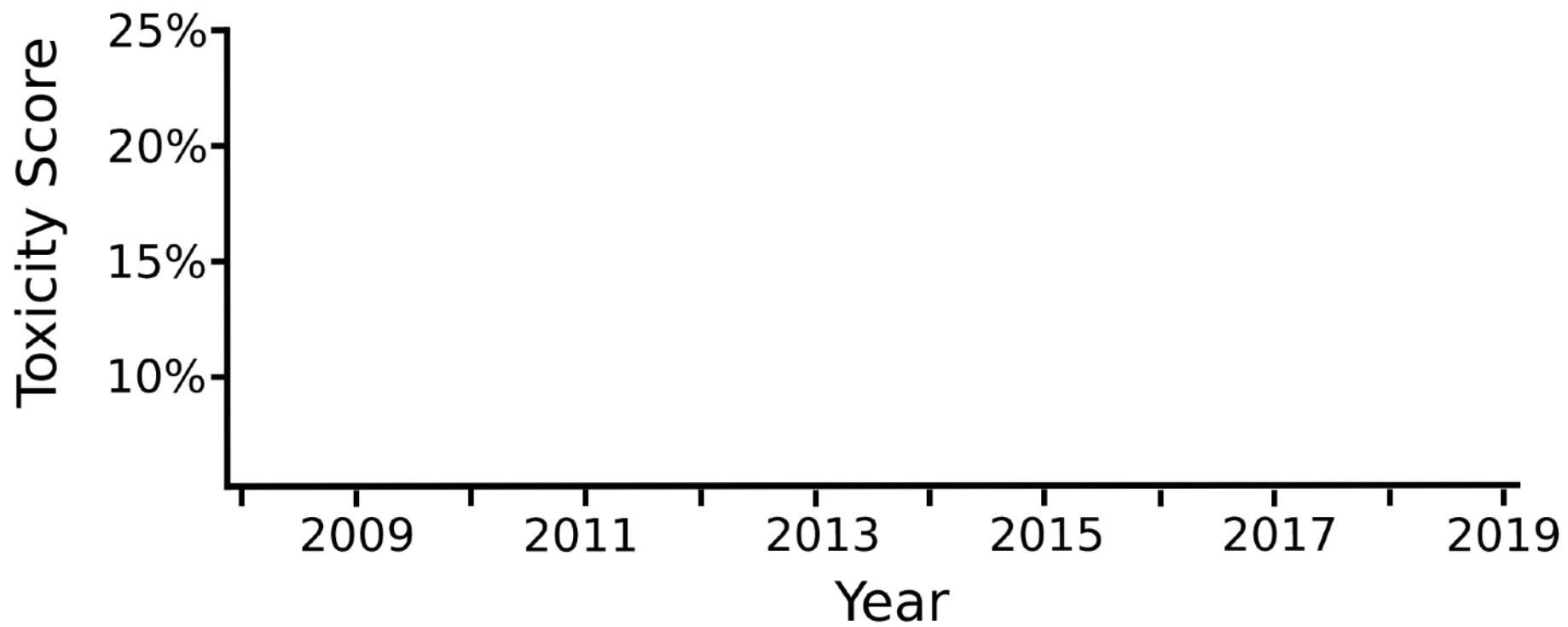
I hope this whore dies.
She has no clue on what it
means to have GAD and panic
attacks.
And I should know...

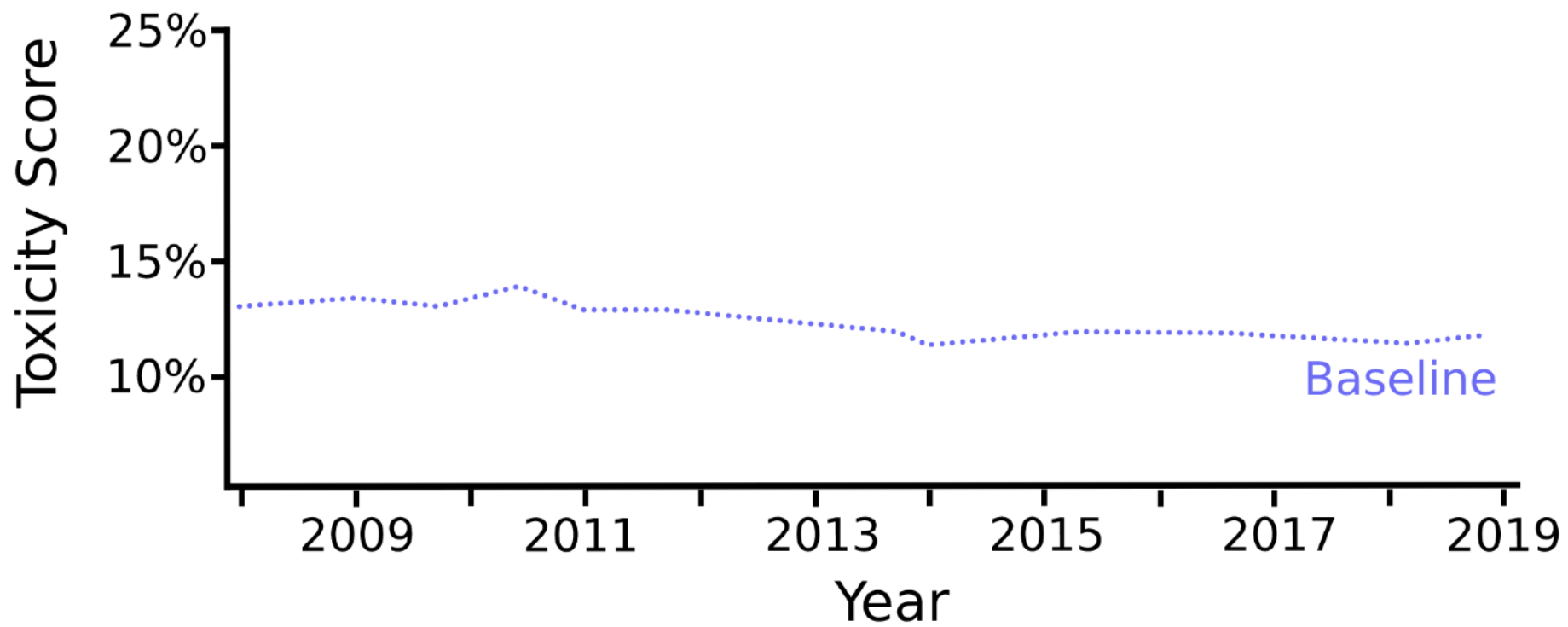


**Perspective
API**

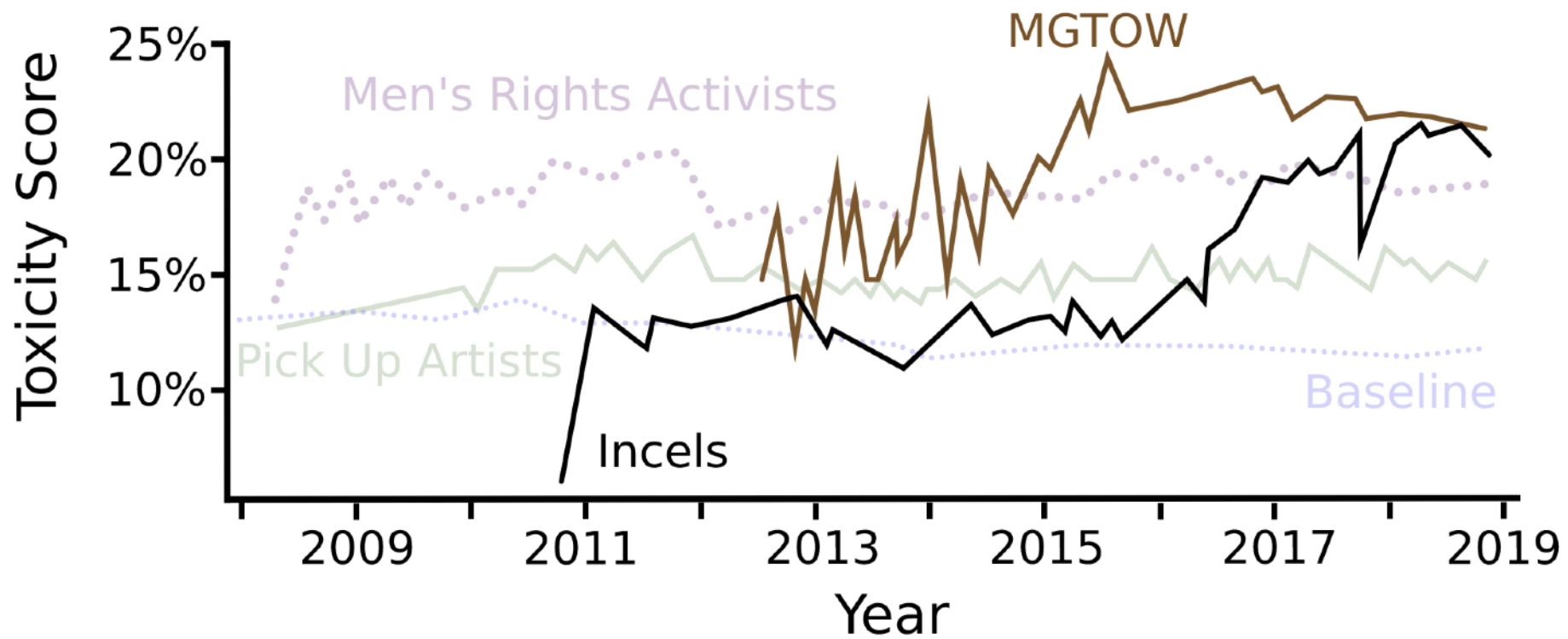


90%





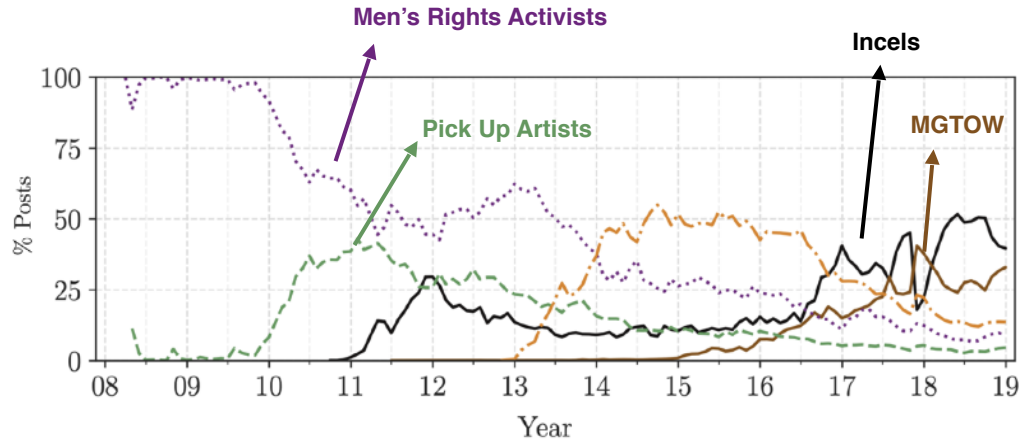




Takeaway #3

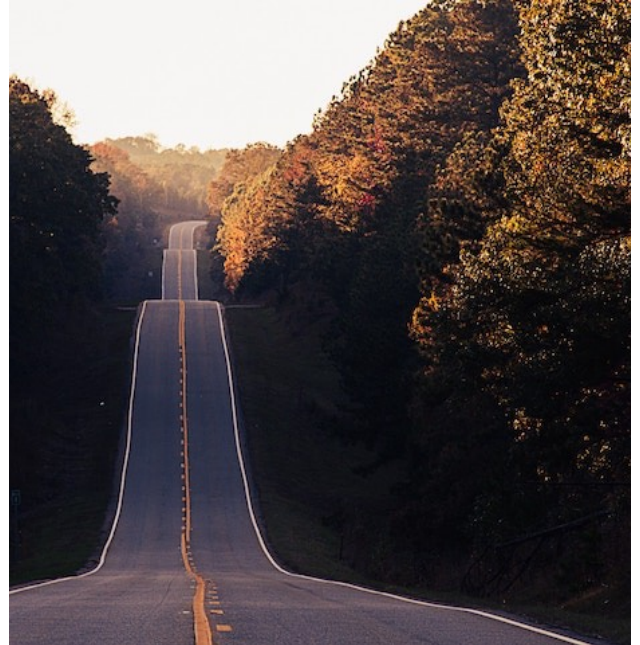
Newer communities such as Incels and MGTOWs are increasingly more toxic.

TL;DR



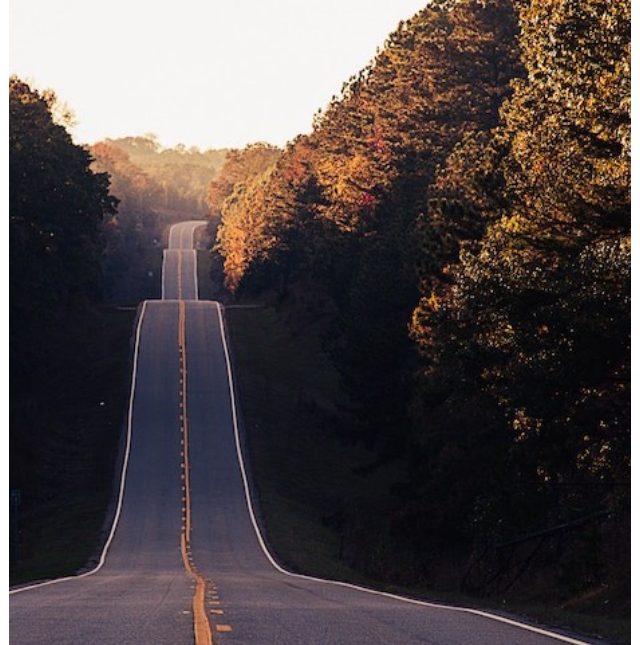
- The Manosphere: Web-based misogynist movements roughly focused on "men's issues," which have seen significant growth over the past years
- Older communities are giving way to more extremist ones
- We observe a substantial migration of active users between communities
- Newer communities are significantly more toxic

Moving Forward



Moving Forward

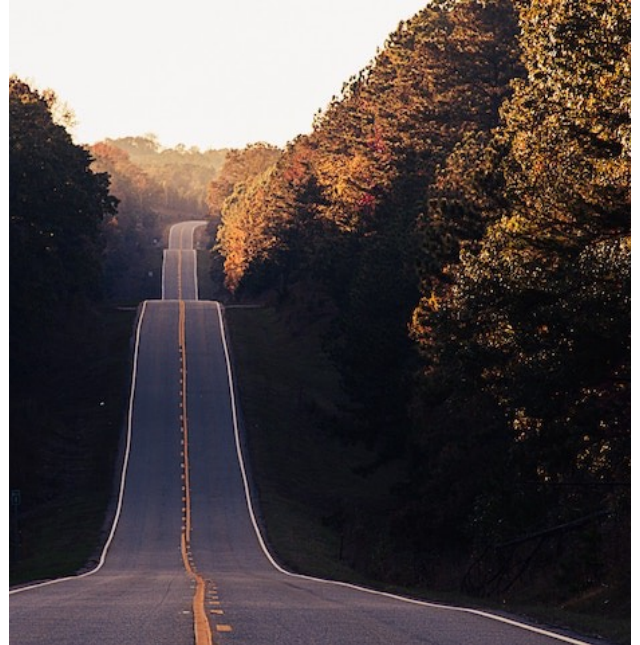
There are still many problems remaining in the broad space



Moving Forward

There are still many problems remaining in the broad space

We have made a lot of progress into gaining a quantitative understanding



Moving Forward

There are still many problems remaining in the broad space

We have made a lot of progress into gaining a quantitative understanding

We now need to deploy **proactive systems in production**

Real time detection

Actively fight weaponized information

Policy

Accountability



Risks

Risks

We've spent a lot of time trying to measure jerks on the Web
They are turning a really powerful technology into a cesspool

Risks

We've spent a lot of time trying to measure jerks on the Web

They are turning a really powerful technology into a cesspool

As researchers, our goal is to share what we learn

We write papers; sometimes people even read them!

Risks

We've spent a lot of time trying to measure jerks on the Web

They are turning a really powerful technology into a cesspool

As researchers, our goal is to share what we learn

We write papers; sometimes people even read them!

Risks

We've spent a lot of time trying to measure jerks on the Web

They are turning a really powerful technology into a cesspool

As researchers, our goal is to share what we learn

We write papers; sometimes people even read them!

Unfortunately, this can attract some unwanted attention...

File: [Cucks.png](#) (209 KB, 1181x535)



The Cuckademics, A Study Anonymous (ID: zw8wNrfR) 10/13/16(Thu)06:18:21 No.92700473

Jeremy Blackburn

>his work on the contagious aspects of cheating and understanding and learning about cyberbullying/toxic behavior in online games represent the largest scale studies of such behavior, providing insight useful not just for the gaming community, but in understanding anti-social behavior in the real-world as well.

<http://www.tid.es/research/researchers/j-blackburn>

Gianluca Stringhini

>Twitter Cuck e.g. "But #brexit has nothing to do with xenophobia, right?"

>Retweets Hillary

https://twitter.com/gianluca_string

Riginos Samaras

>actually seems ok and quite attractive so ill give him a pass

Ilias Leontiadis

>bizarre that they would include this guy because his papers are all related to internet privacy and ad revenue.

>Oh wait... Two factors that may have an influence on such interactions are cultural differences between individuals in highly multicultural workplaces, and the layout and physical spaces of the workplace itself. In this work, we investigate how these two factors may facilitate or hinder inter-group interactions in the workplace.

<http://www.tid.es/research/researchers/ilias-leontiadis>

>> **Anonymous** (ID: zw8wNrfR) 10/13/16(Thu)06:18:40 No.92700514

Nicolas Kourtellis

>Just a henchman

Emiliano De Cristofaro

>globalist

>Italian national Dr Emiliano De Cristofaro, a senior lecturer in computer science at University College London, criticised the comments.

>"Quite frankly, I am very upset. It inherently discriminates against a portion of the population living in this country, making us feel like second-class individuals. Nobody should be used as a negotiating chip, let alone people who are working hard, paying taxes and strongly contributing to the country," he said.

>"I also fear that statements like this will further encourage hate crime, discrimination and racism against EU citizens."

<https://news.co.uk/essentials/news/politics/eu-citizens-living-uk-express-upset-liam-fox-describes-negotiating-card/>

Jeremiah Onaolapo

>Just a henchman but...

>Recent or forthcoming papers: "What Happens After You Are Pwnd: Understanding The Use Of Leaked Webmail Credentials In The Wild."

>In that paper: "We leaked 100 honey accounts on paste sites, underground forums, and virtual machines infected with malware, and provided detailed statistics of the activity of cybercriminals on the accounts, together with a taxonomy of the criminals. Our findings help the research community to get a better understanding of the ecosystem of stolen online accounts, and potentially help researchers to develop better detection systems against this malicious activity."

<http://www0.cs.ucl.ac.uk/staff/J.Onaolapo/papers/imc16gmail.pdf>

Gabriel Emile Hine

>no juice

In conclusion my research indicates that the three most responsible Jews for this vicious hate crime against our culture are:

>Jeremy Blackburn

https://twitter.com/jblackb?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor

>Emiliano De Cristofaro

https://twitter.com/emilianodc_

>Gianluca Stringhini

https://twitter.com/gianluca_string



Shining a light on the dark corners of the web

Cybercrime researcher Gianluca Stringhini explains how he studies hate speech and fake news on the underground network 4chan.

Daniel Cressey

09 June 2017

[Rights & Permissions](#)



Germany focus



The secret to Germany's scientific excellence

With a national election this month, Germany proves that foresight and stability can power research.

Recent

Read

Commented

1. [Deadly Mexico earthquake had unusual cause](#)
Nature | 08 September 2017
2. [Geneticists pan paper that claims to predict a person's face from their DNA](#)
Nature | 08 September 2017
3. [South Korean researchers lobby government to lift human-embryo restrictions](#)
Nature | 08 September 2017





Shining a light on the dark corners of the web

Cybercrime researcher Gianluca Stringhini explains how he studies hate speech and fake

Germany focus



ic excellence

Germany proves
ar research.

Commented

ad unusual

laims to predict
NA

Nature | 06 September 2017

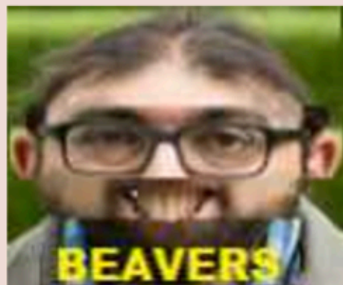
3. South Korean researchers lobby
government to lift human-embryo
restrictions

Nature | 08 September 2017

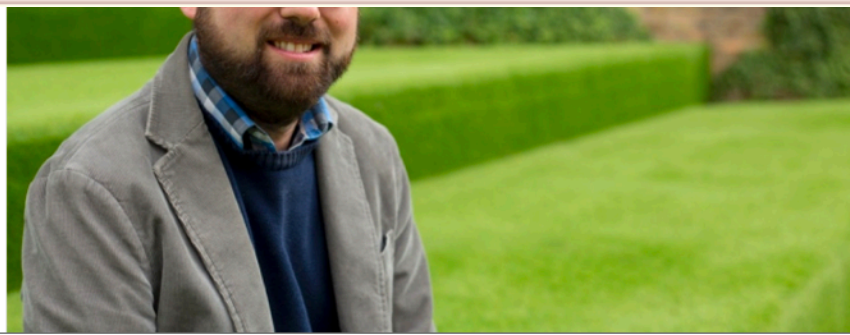


☐ Anonymous (ID: 0a1kdjN4)  06/10/17(Sat)02:11:32 No.129265175 ► >>129270300

File: [BEAVERRS.png](#) (26 KB, 126x105)



[>>129263649](#)



nature

Home | News & Comm

News & Comment > N

NATURE | NEWS: Q&A

Shining a light

Cybercrime research

hello i am
italian

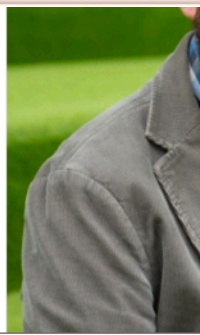
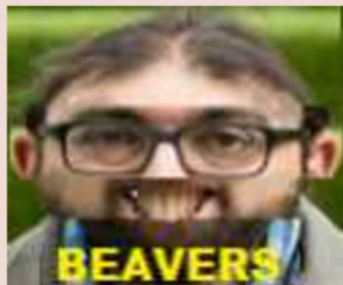


MOSHE P.I.

***The Cyberhate
Investigator***

☐ Anonymous (ID: 0alkdjN4

File: [BEAVERRS.png](#) (26 KB,



nature
briefing



nature

Home | News & Comm

News & Comment > N

NATURE | NEWS: Q&A

Shining a light

Cybercrime research

hello i am
italian



MOSHE P.I.

***The Cyberhate
Investigator***

☐ **Anonymous** (ID: 0a1kdjN4)
File: [BEAVERSong](#) (26 KB)



From what we could
determine, this image
was produced by Daily
Stormer users...

A literal, non-satirical
neo-Nazi community

nature
briefing



Provided inspiration for new memes



Figure 29: A very comfy Savvas Zannettou

Thank you!

