

# Understanding and Countering the Weaponization of the Web

Emiliano De Cristofaro  
[emilianodc.com](http://emilianodc.com)



**IDRAMA LAB**





**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

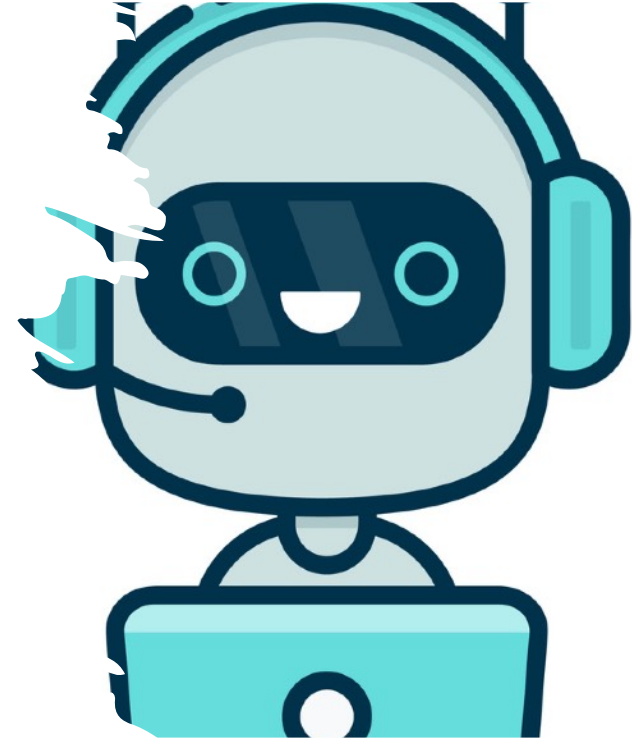
Election  
Interference

# What can we do?



# What can we do?

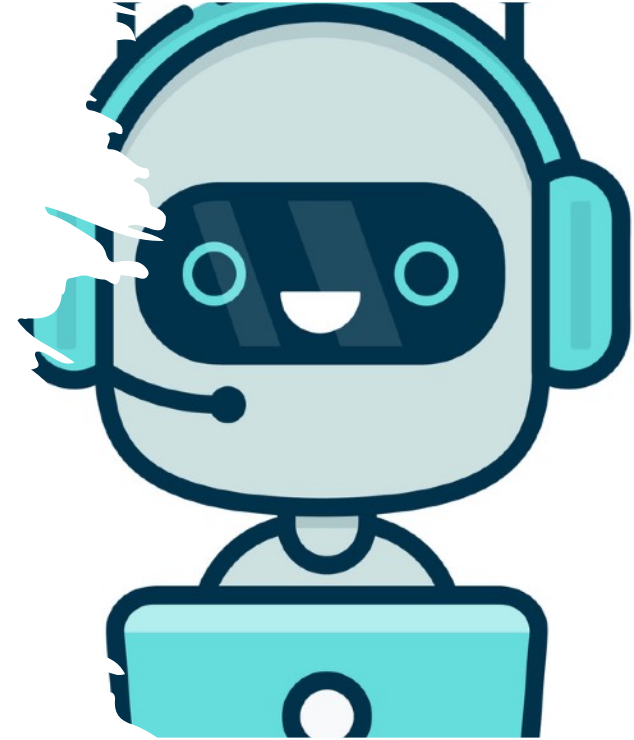
We have experience with **mitigating unwanted activity** on social networks!



# What can we do?

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

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

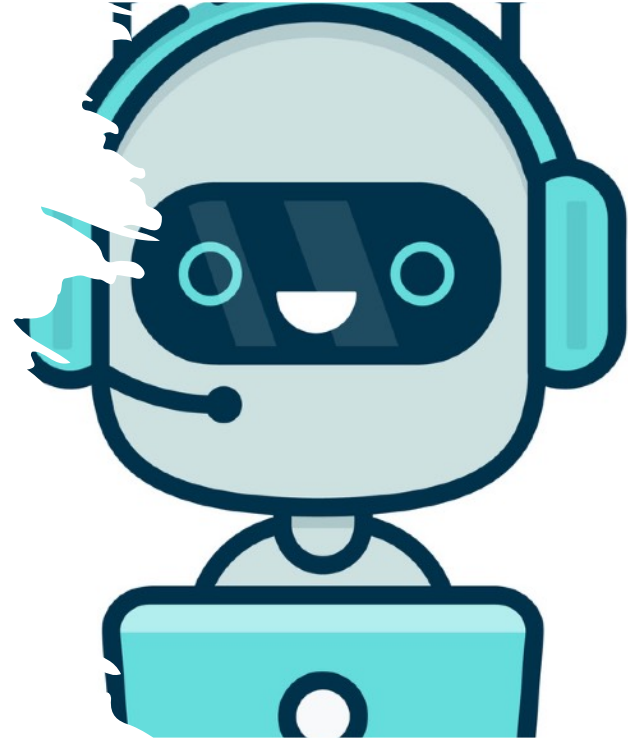


# What can we do?

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.

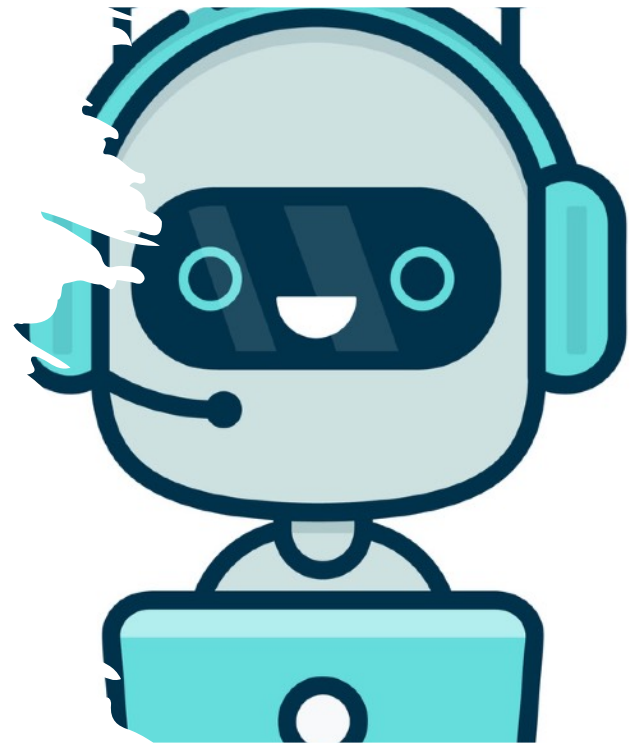


# What can we do?

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.

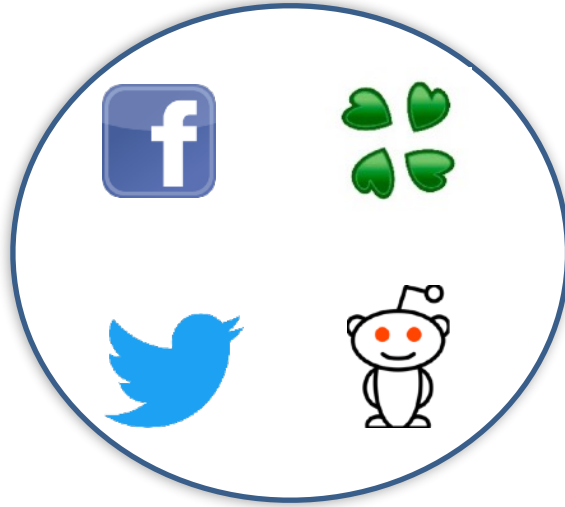
And, we can't take the fight on merely **economic grounds**

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



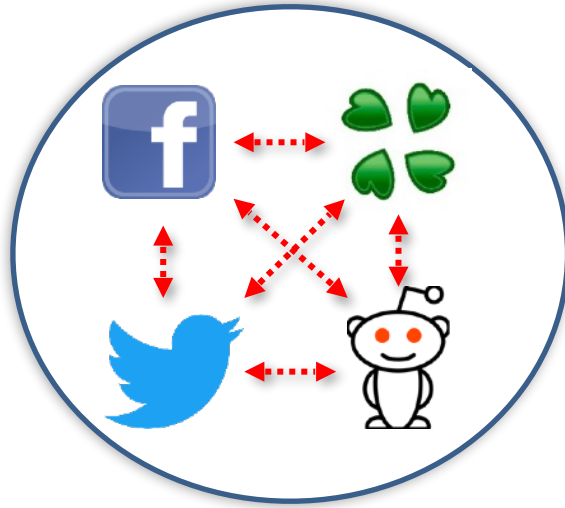
Online services do not exist in a vacuum

# Online services do not exist in a vacuum

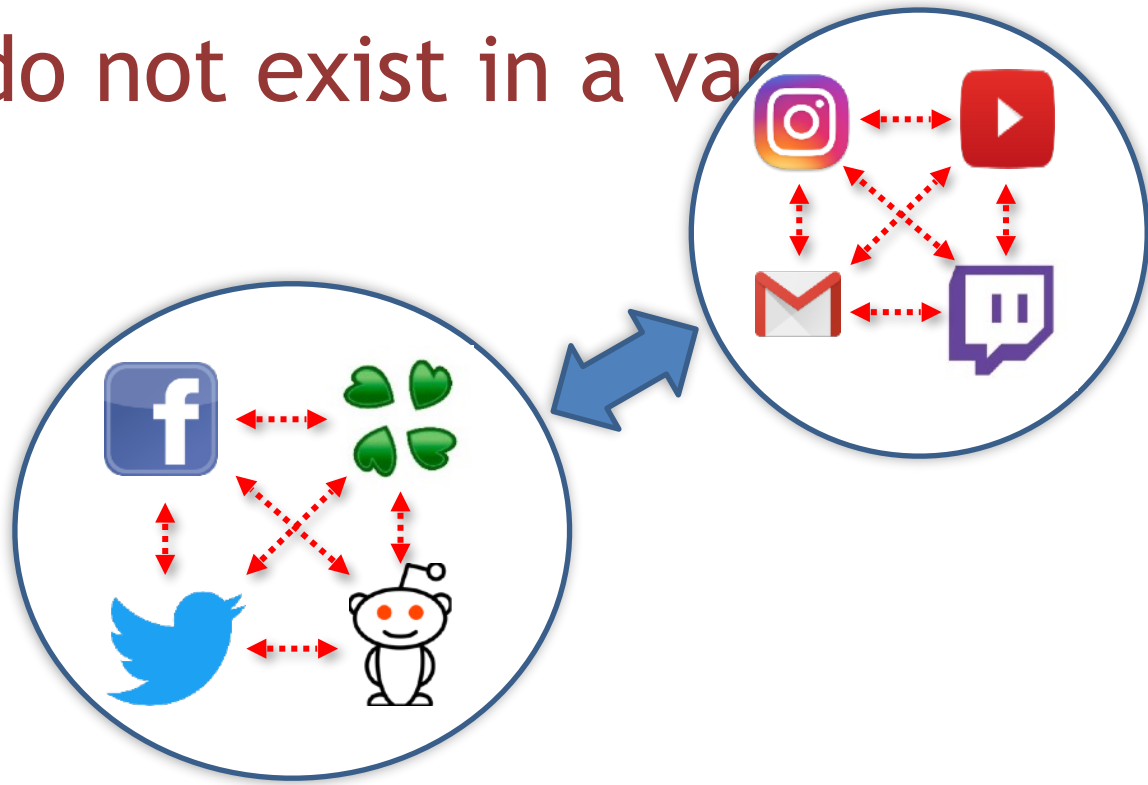




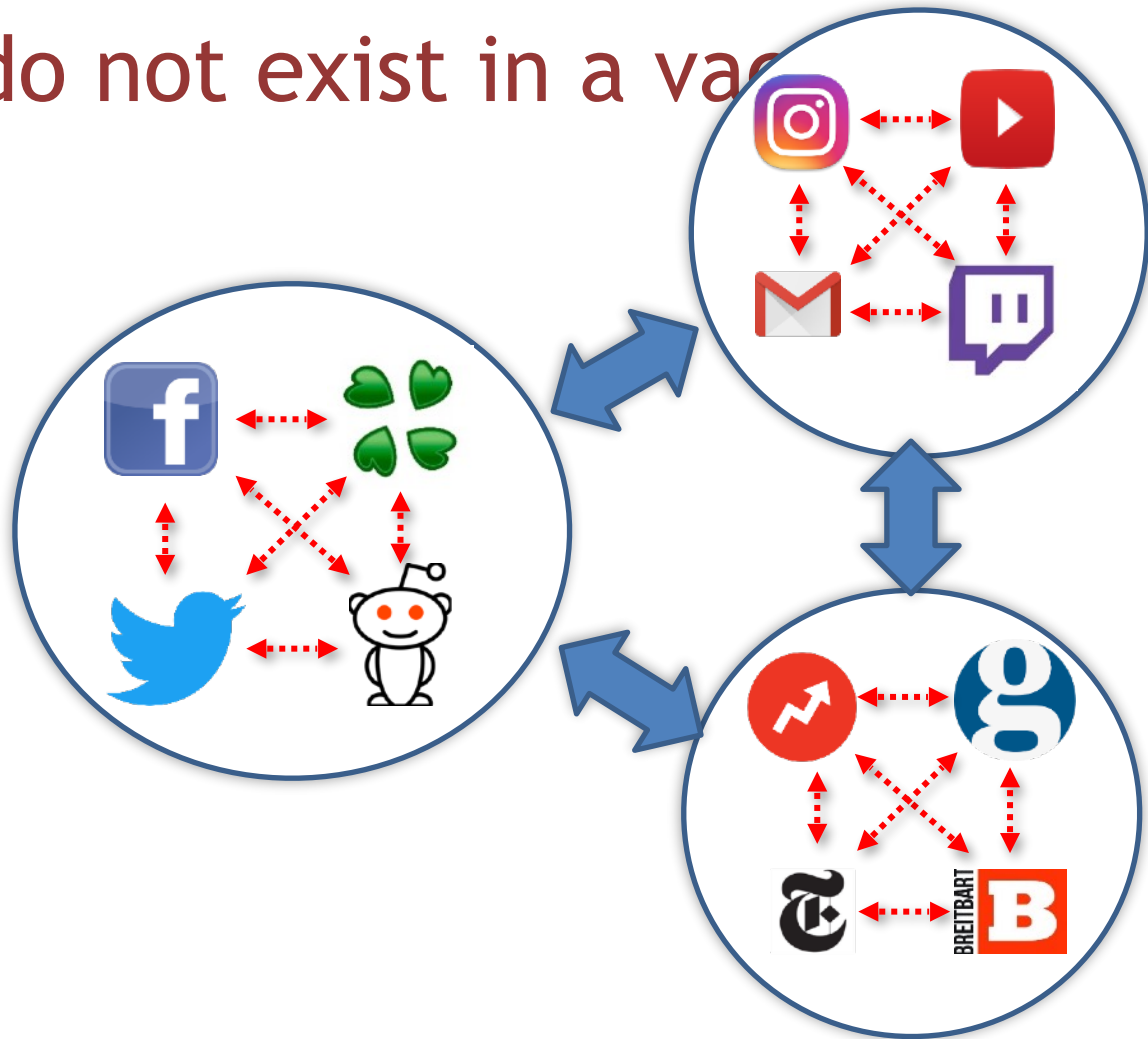
# Online services do not exist in a vacuum



Online services do not exist in a vacuum

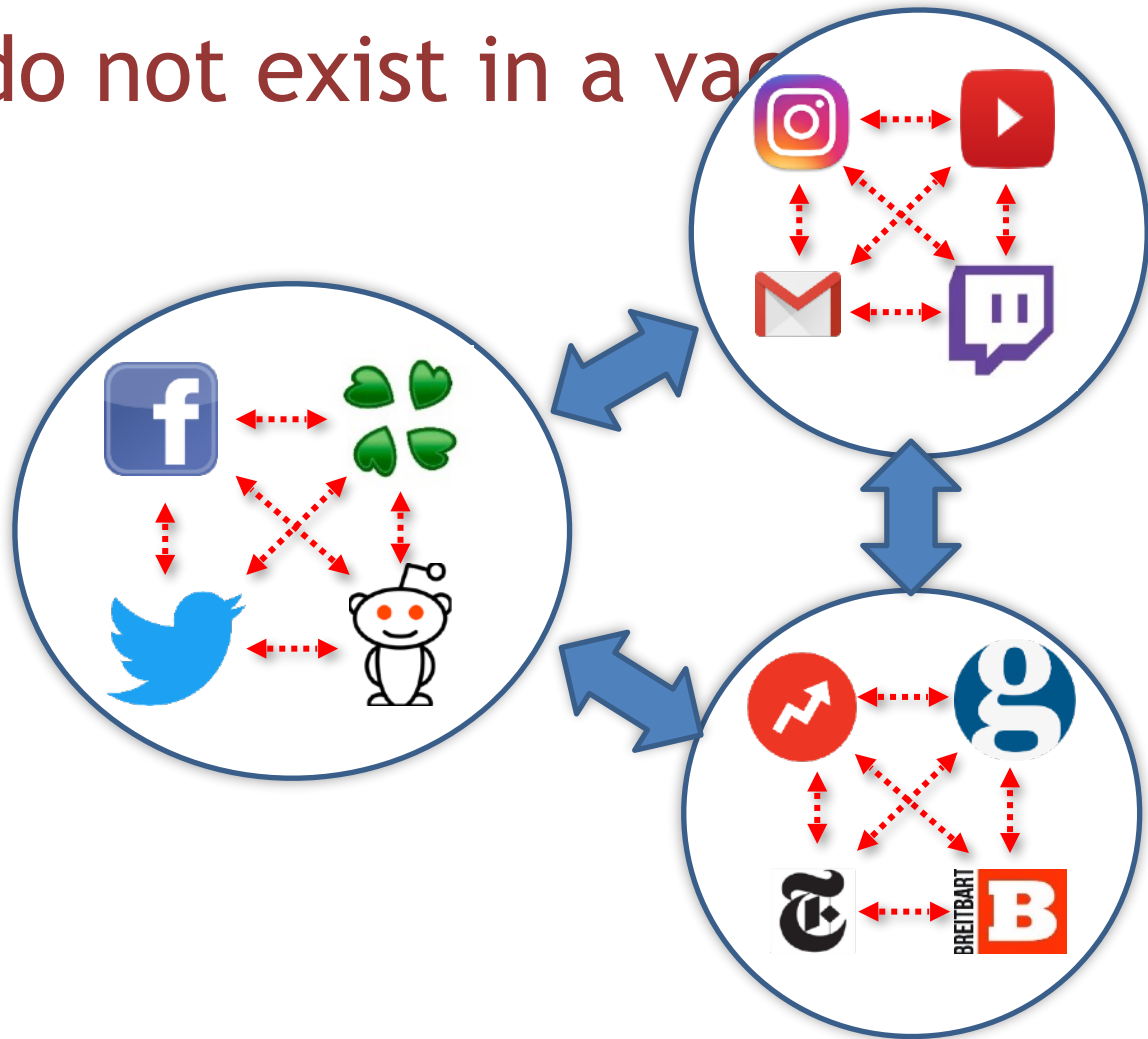


Online services do not exist in a vacuum



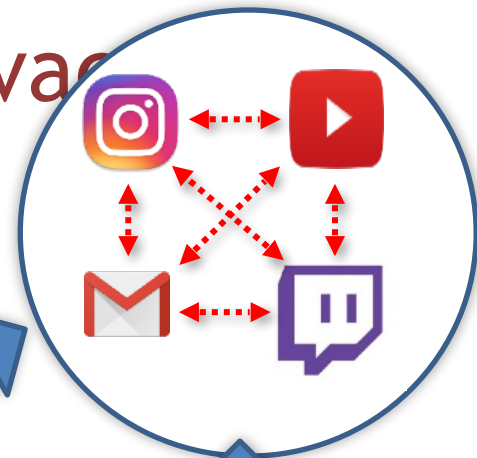
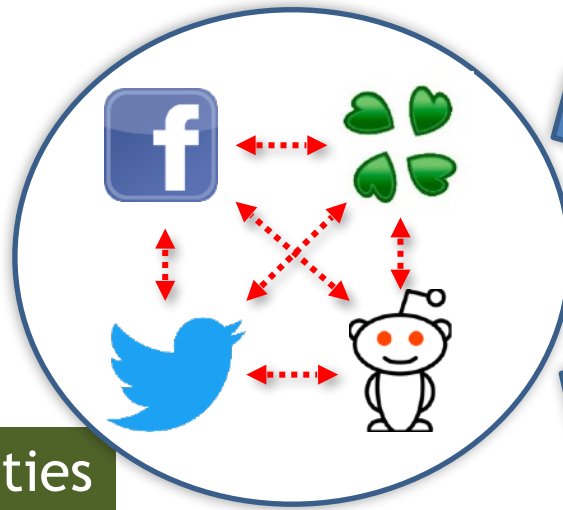
# Online services do not exist in a vacuum

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

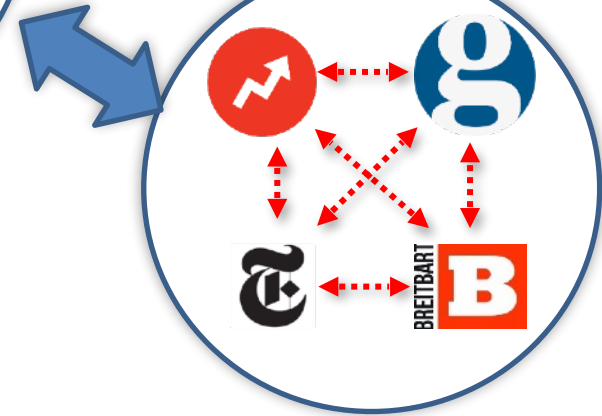


# Online services do not exist in a vacuum

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

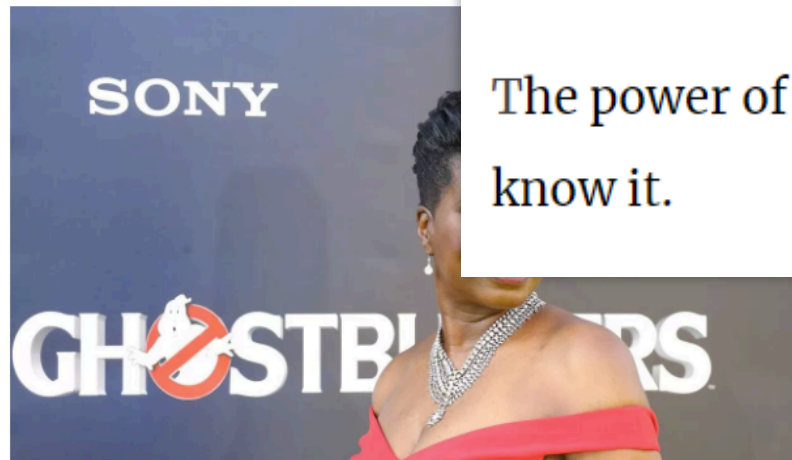
# Pizzagate: How a 4Chan conspiracy went mainstream

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



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?**

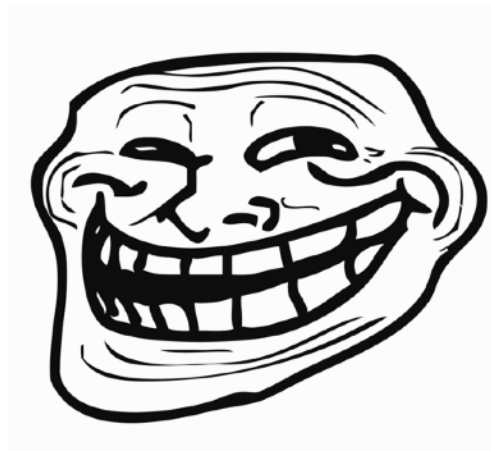




He



He





He



He

I CAN HAS  
CHEEZBURGER?



Donald J. Trump  
@realDonaldTrump

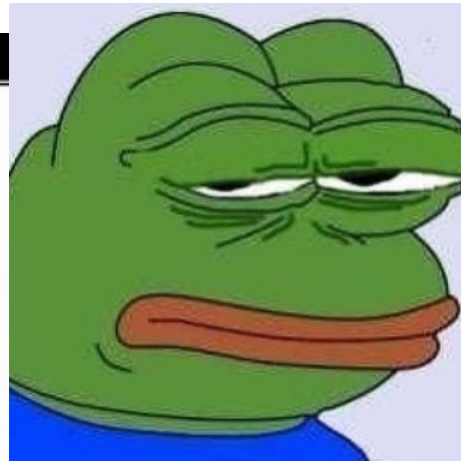
Follow

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



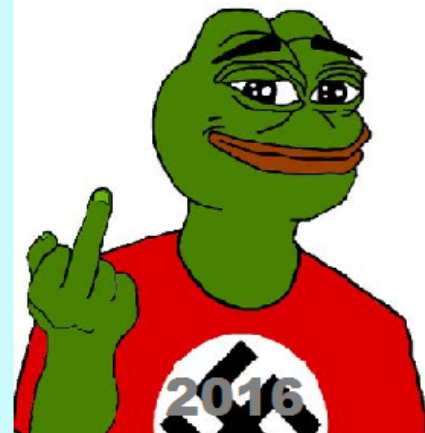
ANONYMOUS

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



He

I CAN HAS  
CHEEZBURGER?



**WHERE DID EVERYTHING  
GO SO REICH**



# 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 at a time

Other users can

With or without

etc.

Boards					filter ▼
<u><b>Japanese Culture</b></u>	<u><b>Interests</b></u>	<u><b>Creative</b></u>	<u><b>Other</b></u>	<u><b>Adult (NSFW)</b></u>	
Anime & Manga	Comics & Cartoons	Oekaki	Business & Finance	Sexy Beautiful Women	
Anime/Cute	Technology	Papercraft & Origami	Travel	Hardcore	
Anime/Wallpapers	Television & Film	Photography	Fitness	Handsome Men	
Mecha	Weapons	Food & Cooking	Paranormal	Hentai	
Cosplay & EGL	Auto	Artwork/Critique	Advice	Ecchi	
Cute/Male	Animals & Nature	Wallpapers/General	LGBT	Yuri	
Flash	Traditional Games	Literature	Pony	Hentai/Alternative	
Transportation	Sports	Music	Current News	Yaoi	
Otaku Culture	Alternative Sports	Fashion	Worksafe Requests	Torrents	
<u><b>Video Games</b></u>	Science & Math	3DCG	Very Important Posts	High Resolution	
Video Games	History & Humanities	Graphic Design	<u><b>Misc. (NSFW)</b></u>	Adult GIF	
Video Game Generals	International	Do-It-Yourself	Random	Adult Cartoons	
Pokémon	Outdoors	Worksafe GIF	ROBOT9001	Adult Requests	
Retro Games	Toys	Quests	Politically Incorrect		
			International/Random		
			Cams & Meetups		
			Shit 4chan Says		





/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)



This  
issue

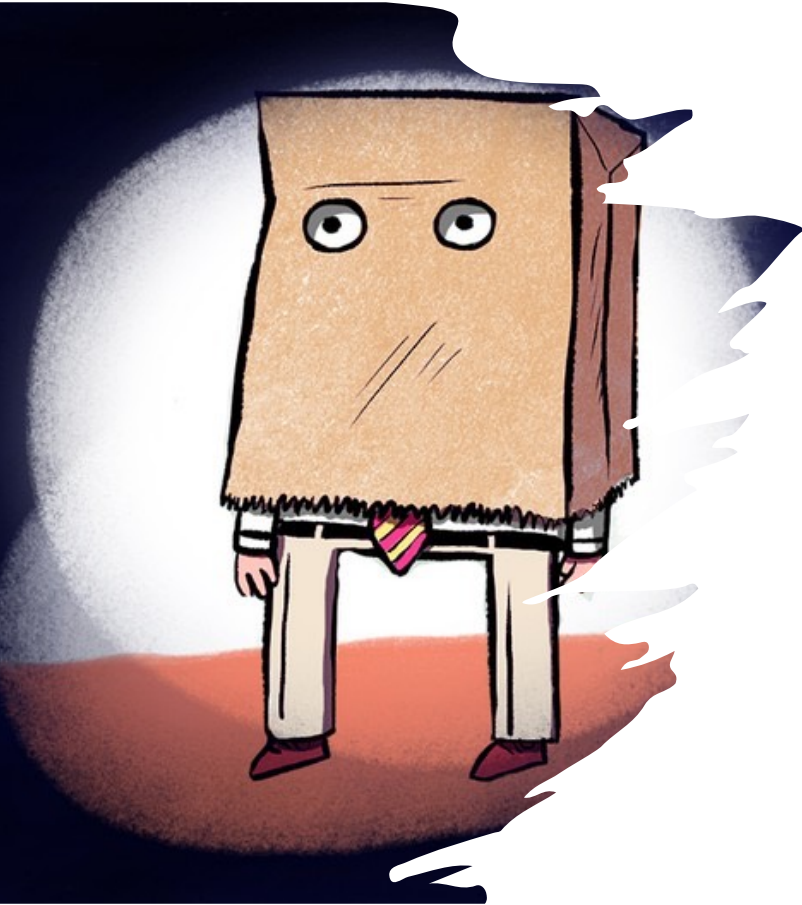
Extremely lax moderation

Almost anything goes

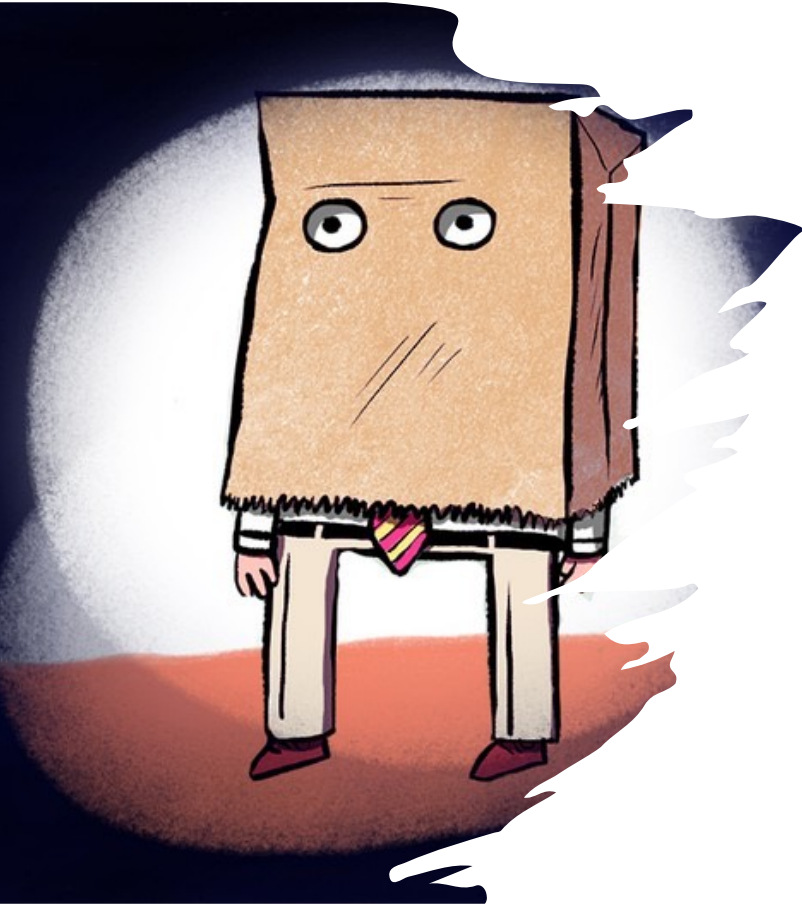
- >Red pill me on X. (with no extra
- >Are X white?
- >Is X degeneracy?
- >How come X girls love Y guys so much?
- >If X is true, then how come Y? Check

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



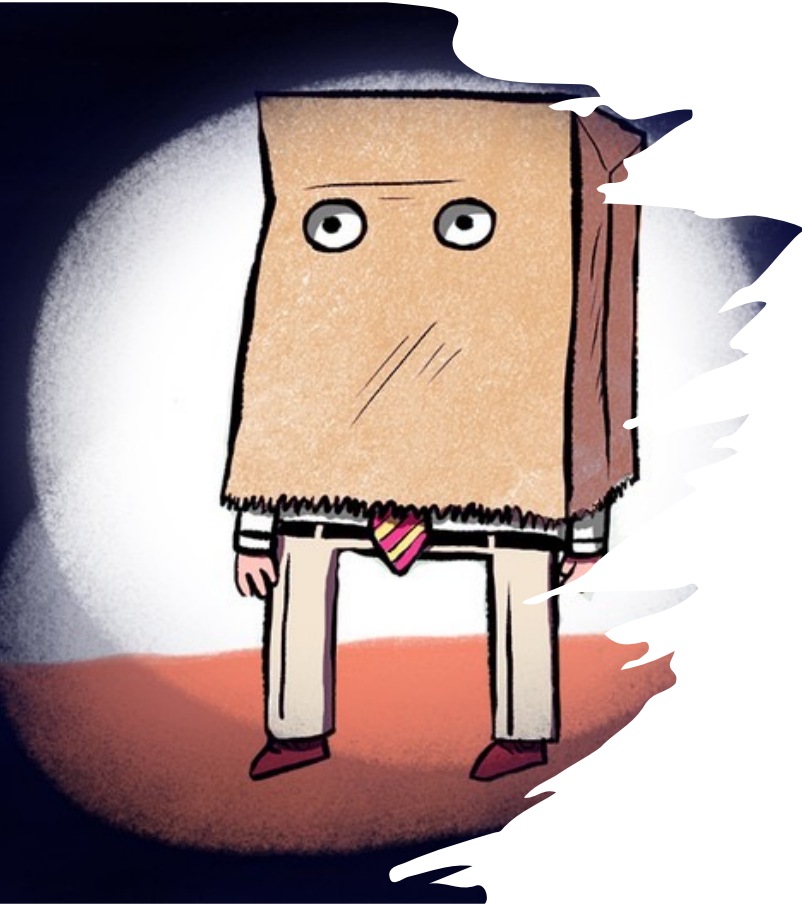
# Anonymity & Ephemerality



**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**

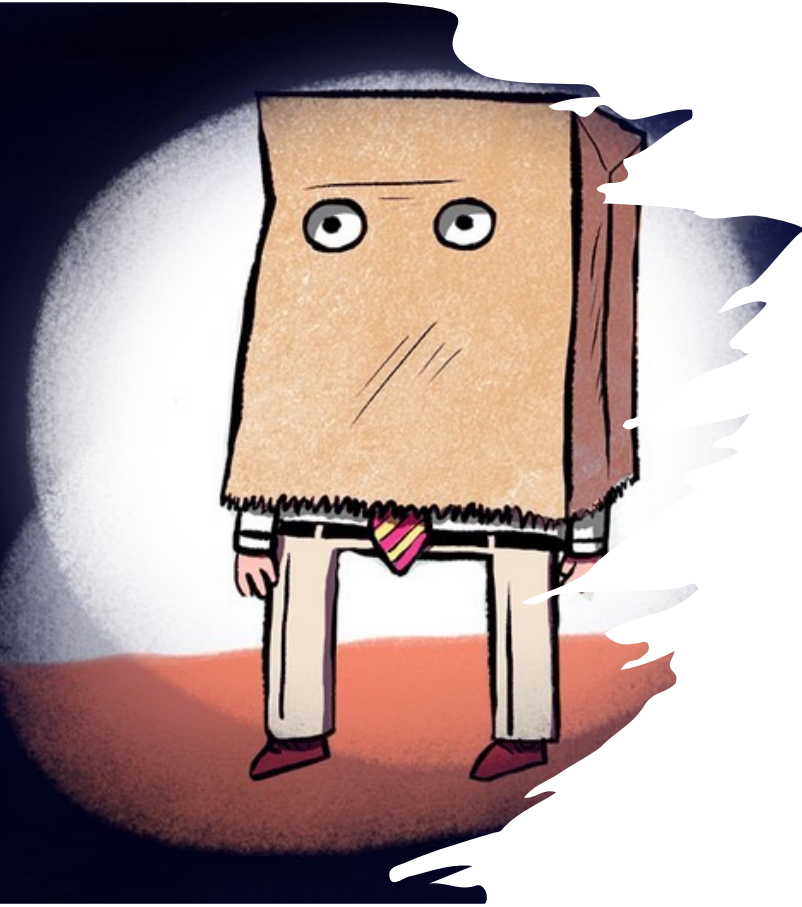
Anonymity is the default (and preferred) behavior

**“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

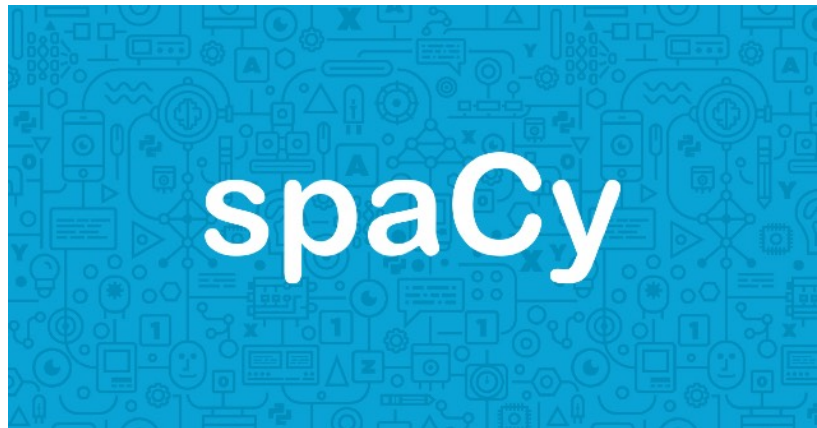
~3.4 million threads from /pol/

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

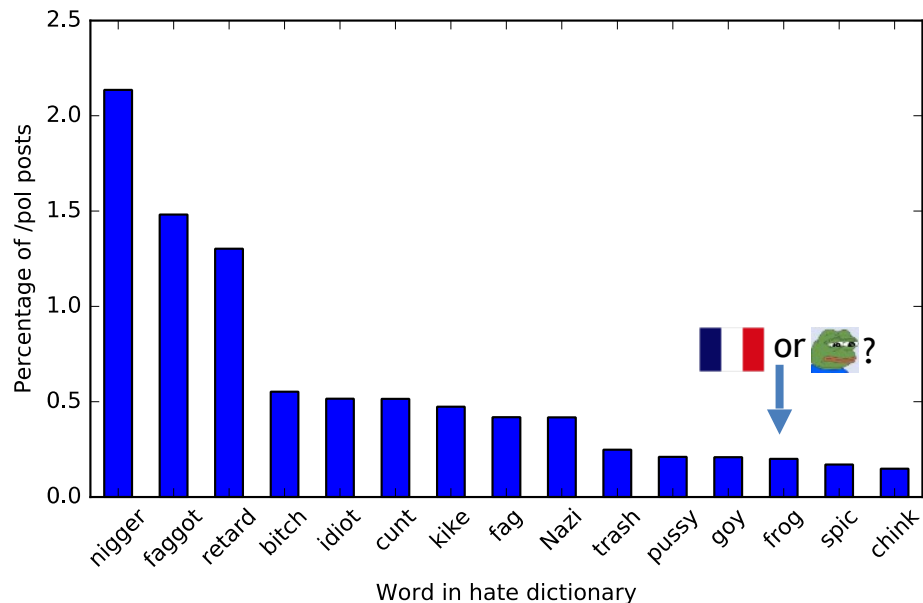
/pol/ by far most  
hate speech use

/pol/ 12%

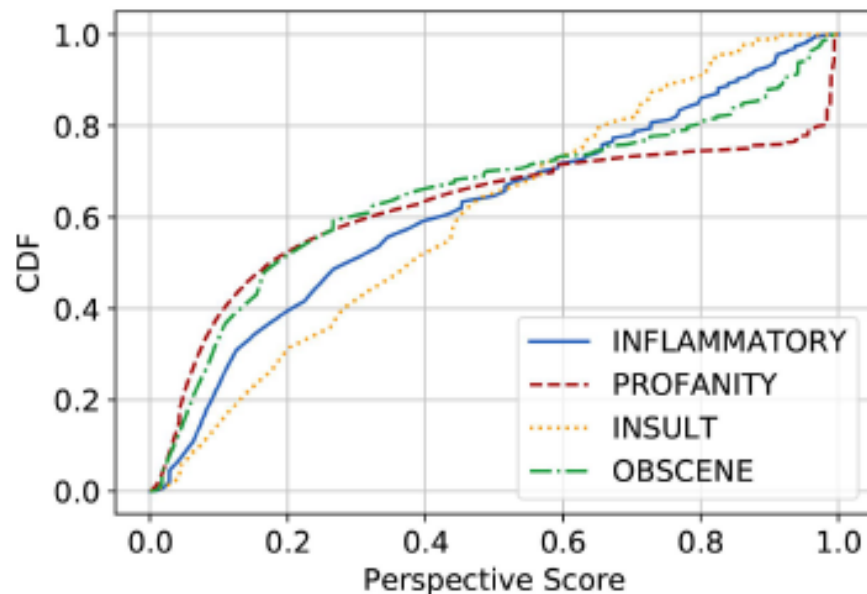
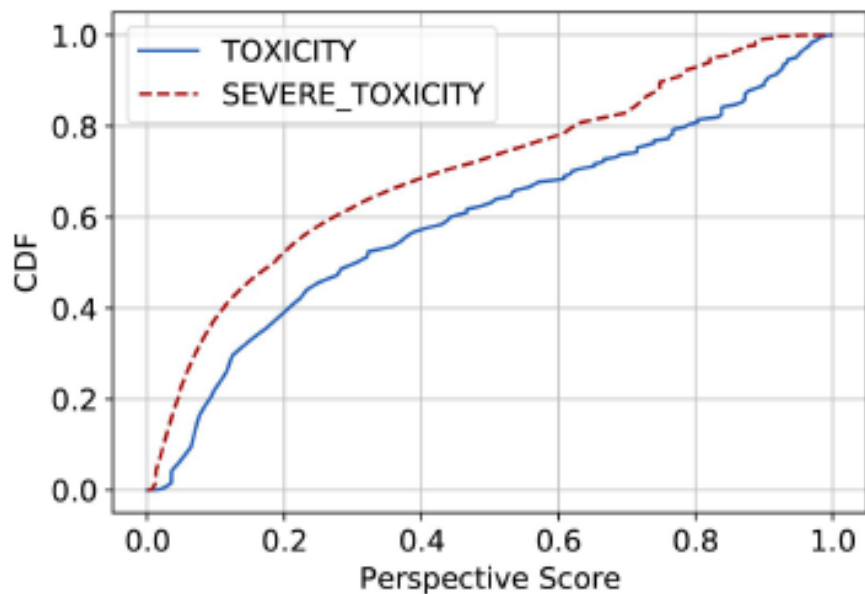
/sp/ 7.3%

/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”)

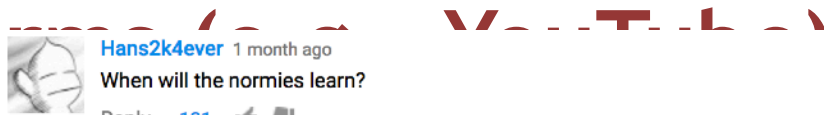
Someone posts a YouTube link  
Maybe with a prompt like “you  
know what to do”  
YouTube video starts getting  
hateful comments

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 platfo



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

Emiliano De Cristofaro

11,013 views

+ Add to Share \*\*\* More

80 526

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

Sz Szgerle 1 month ago

Not even just your regular nazi but ULTRA NAZI!

Reply • 1

Drakewood 1 month ago

+Potato Jenova A FUCKING LEAF!

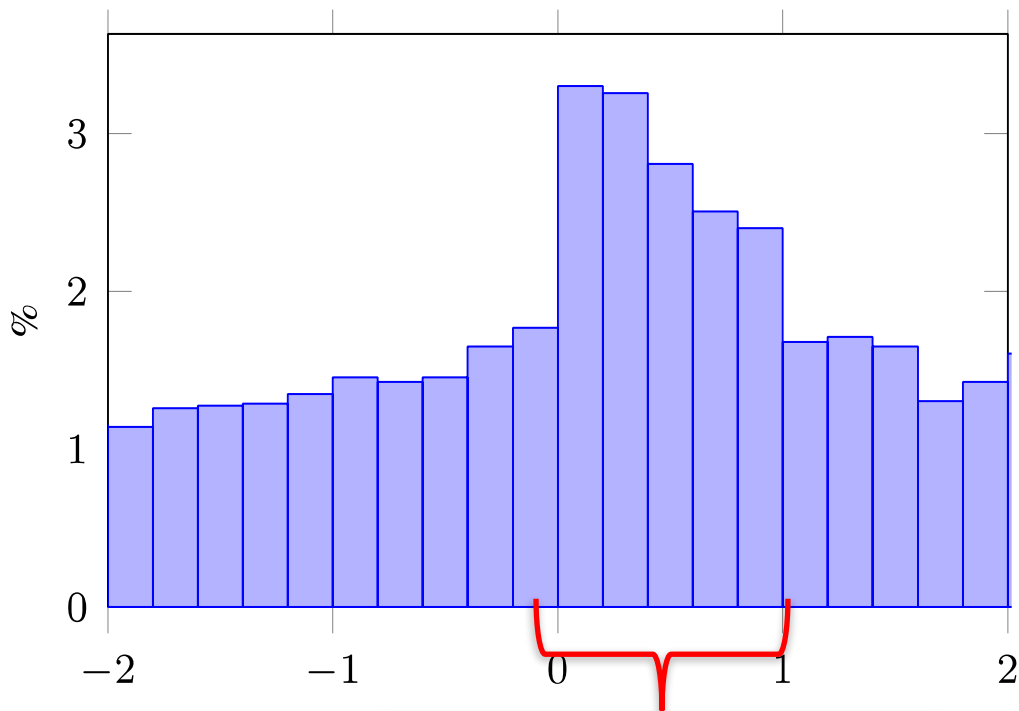
Reply •

Thought Criminal 1 month ago

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

Reply • 169

# Activity Peaks



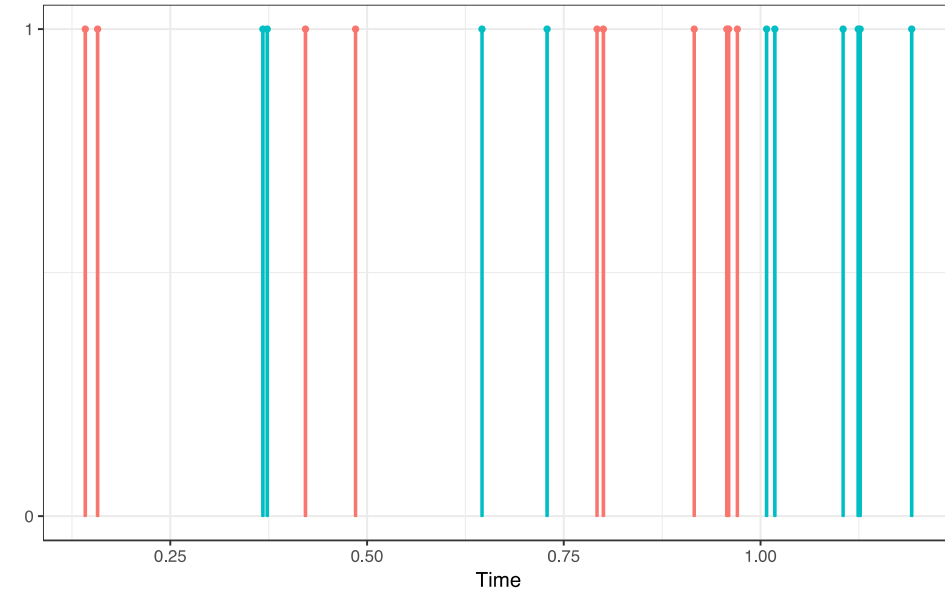
YT videos with peaks  
during 4chan thread

Determined via PDF of  
commenting timeseries

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

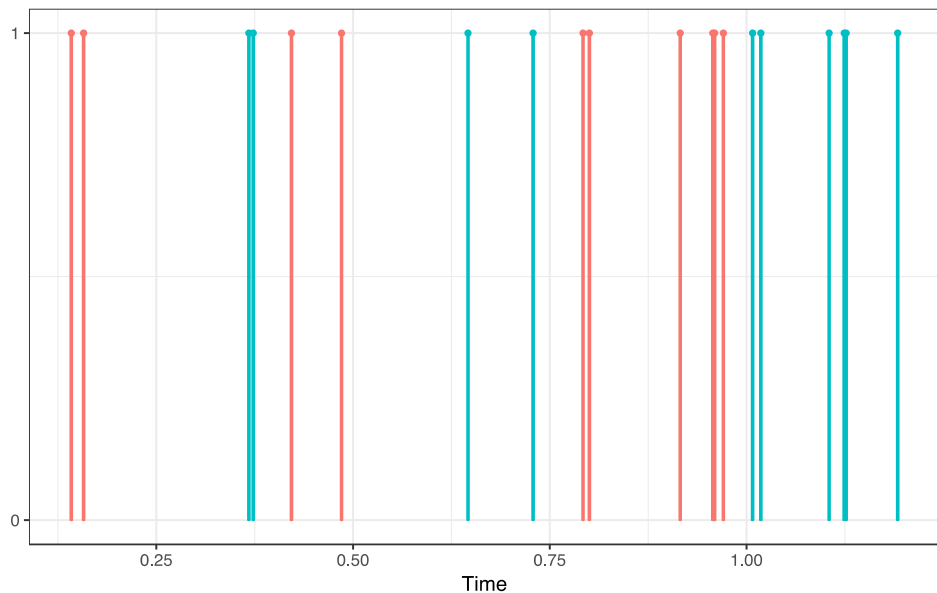
# Synchronization

# Synchronization

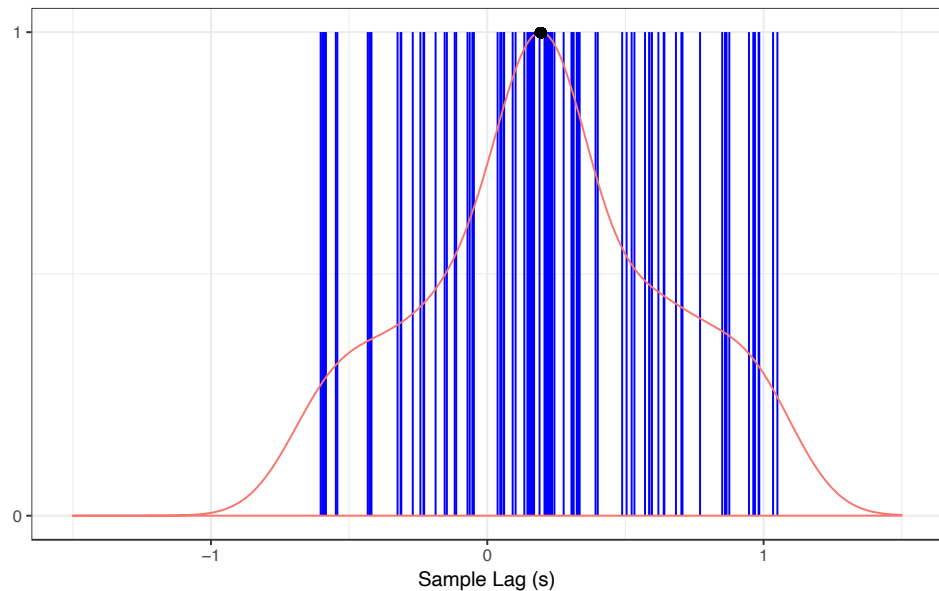


Two series, second randomly shifted from first by 0.2s on avg

# Synchronization



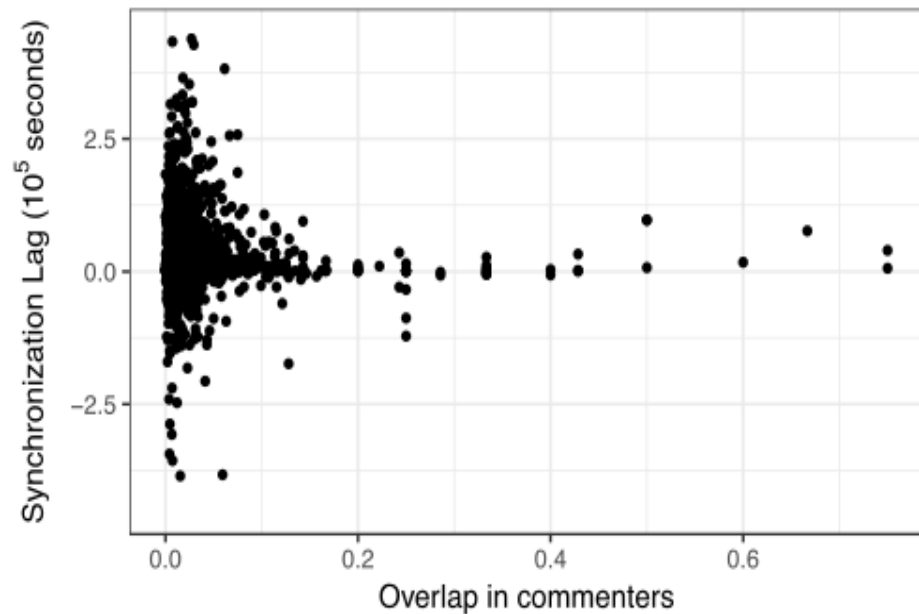
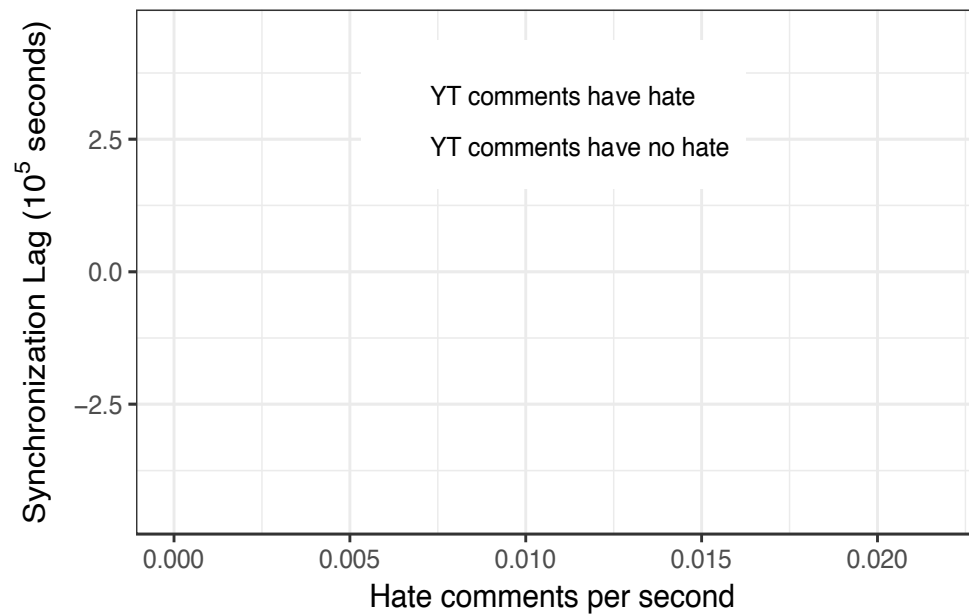
Two series, second randomly shifted from first by 0.2s on avg



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

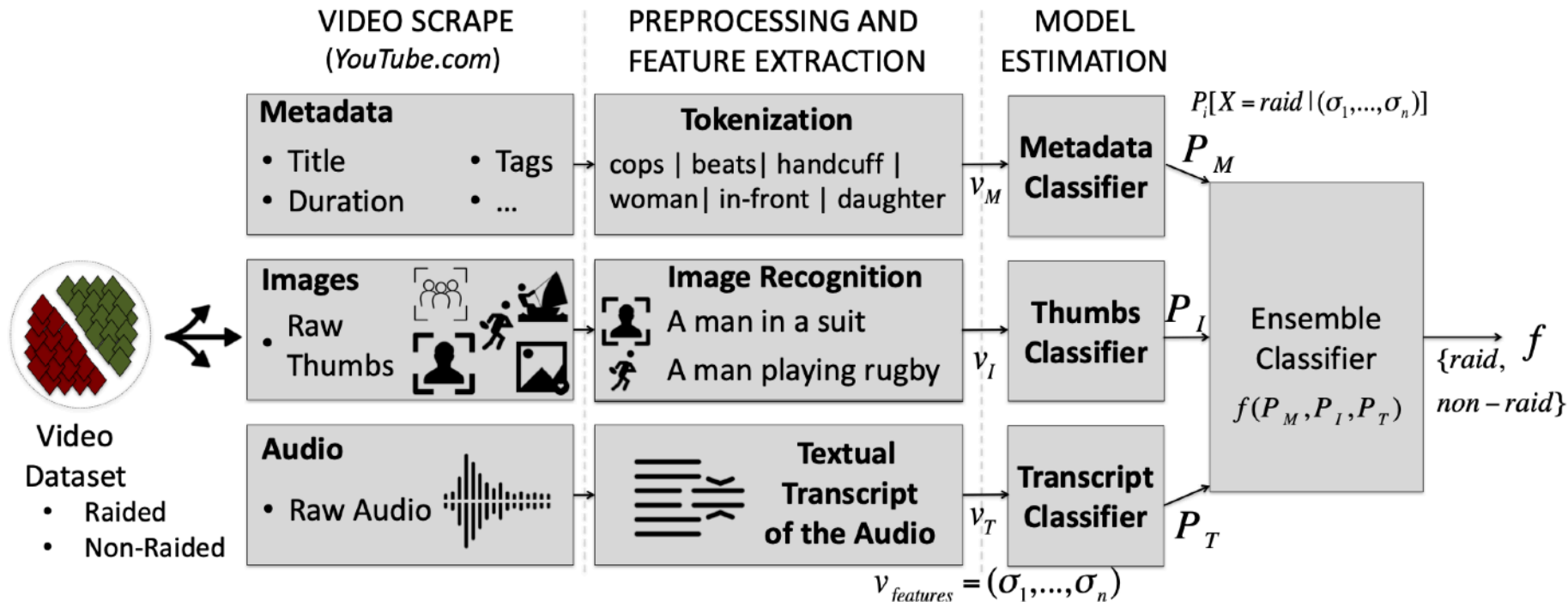
Moderators going through the flagged content

They rarely use fully automated checks to avoid FP

Inherently slow

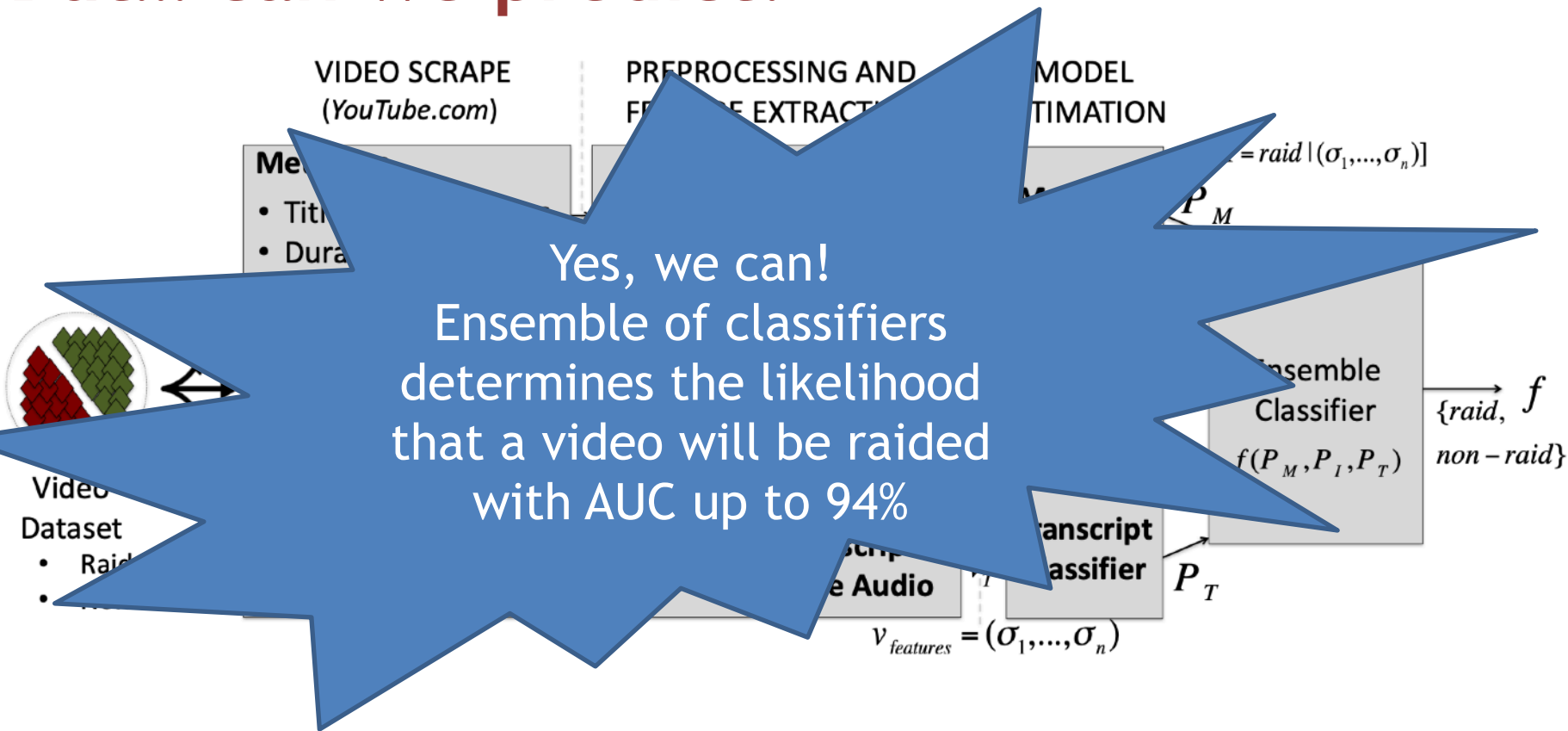


# But... can we predict?

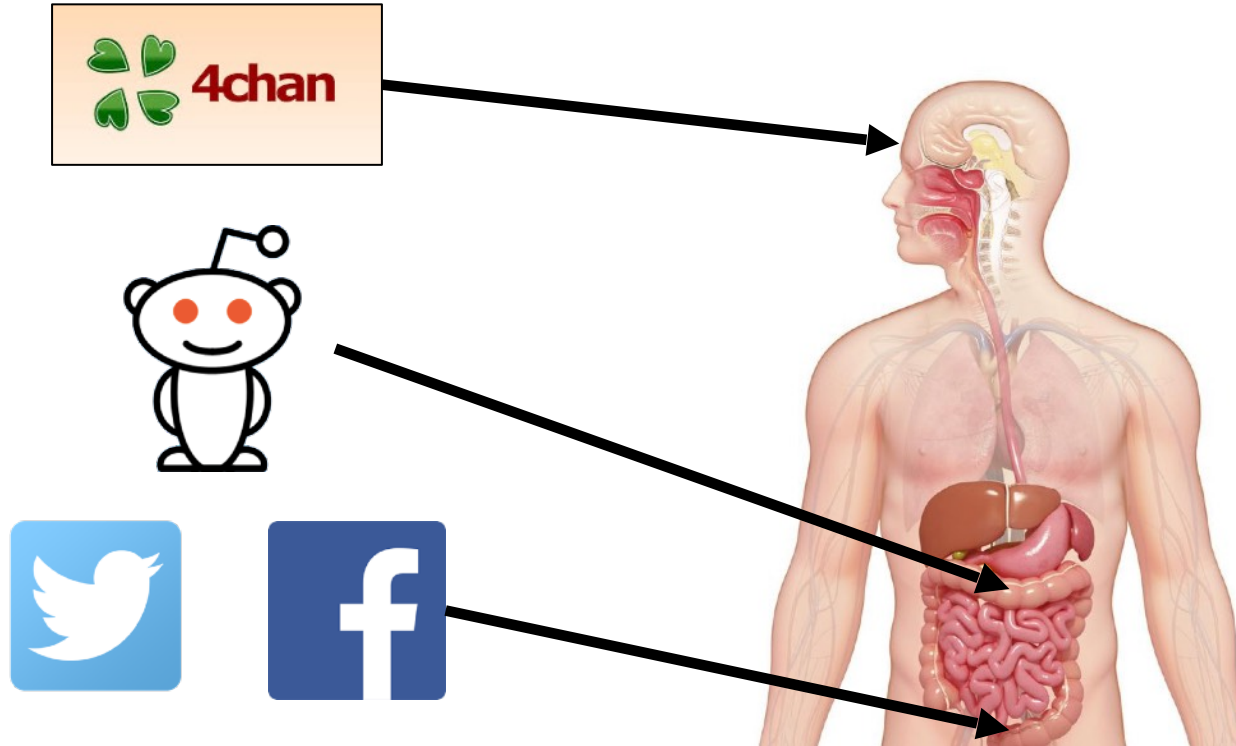




# But... can we predict?



# 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



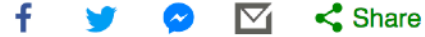
# The Pizzagate Conspiracy Theory

**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**

# 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...

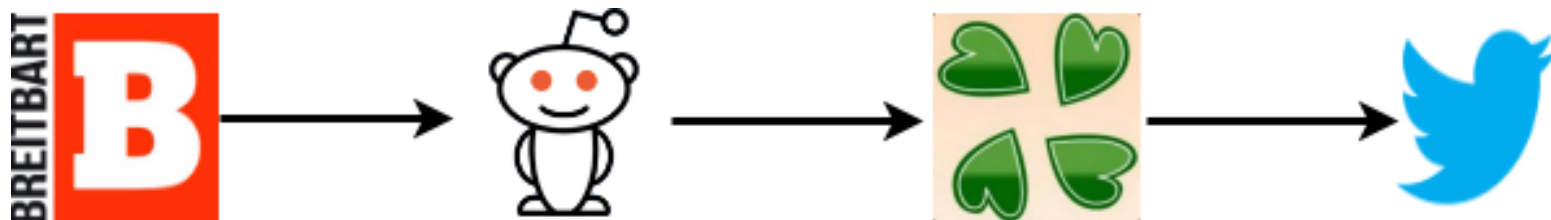
Study the appearance of alternative and mainstream news URLs within the platforms

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

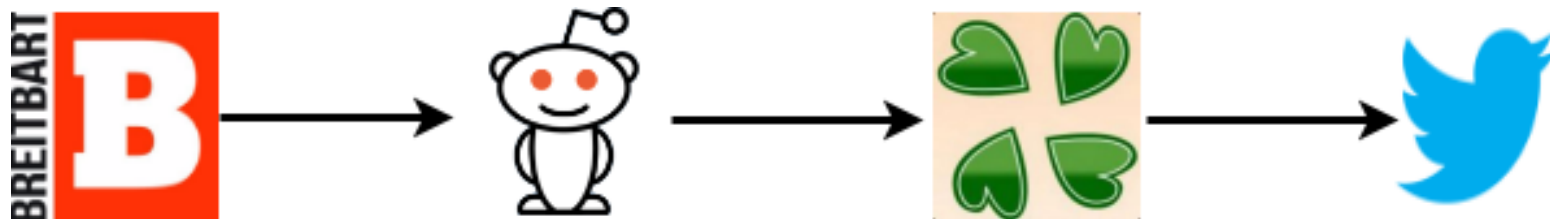
Build a sequence of appearance for each URL according to the timestamps



# 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



Build a graph with the sequences

# 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%)

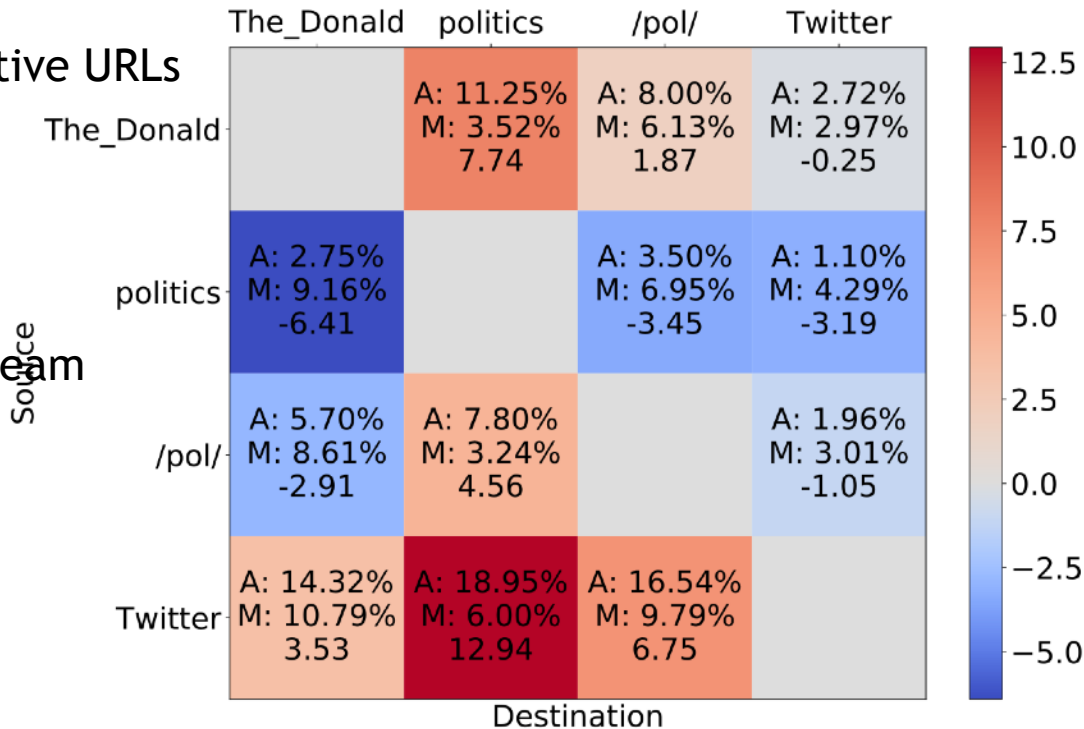
# How Communities Influence Each Other?

## 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?

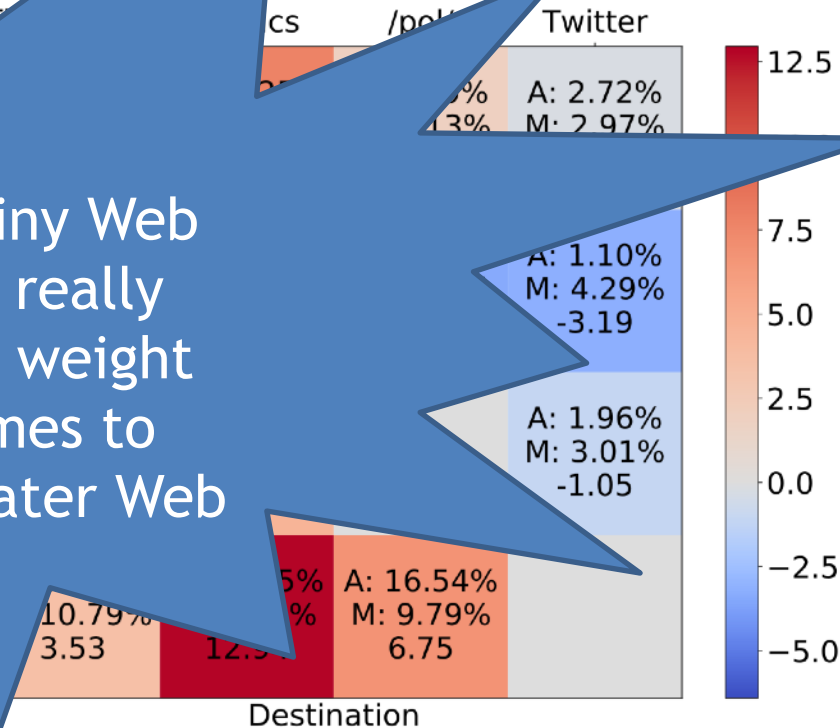
Twitter top influencers for alternative media

- /r/The\_Donald
- /pol/ (1.96%)

Twitter top influencers for LIBERAL

- /r/

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

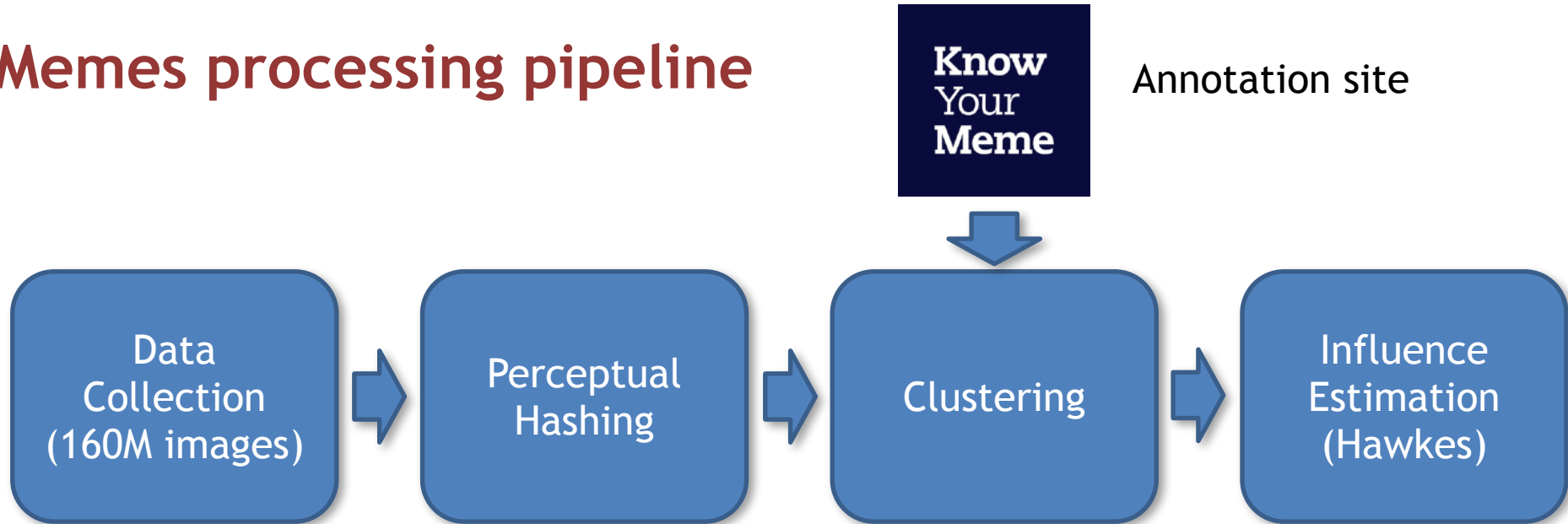
ist: Trump  
young black leaders  
wheeling rant

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](https://github.com/memespaper/memes_pipeline)

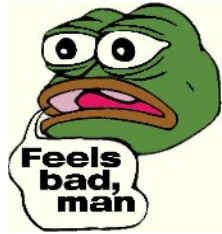


# Examples of clusters

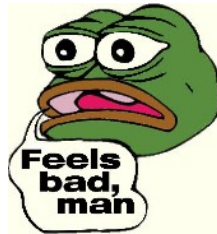


# Top memes per social network

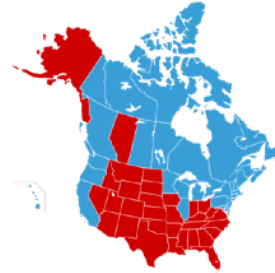
4chan (/pol/)



Reddit

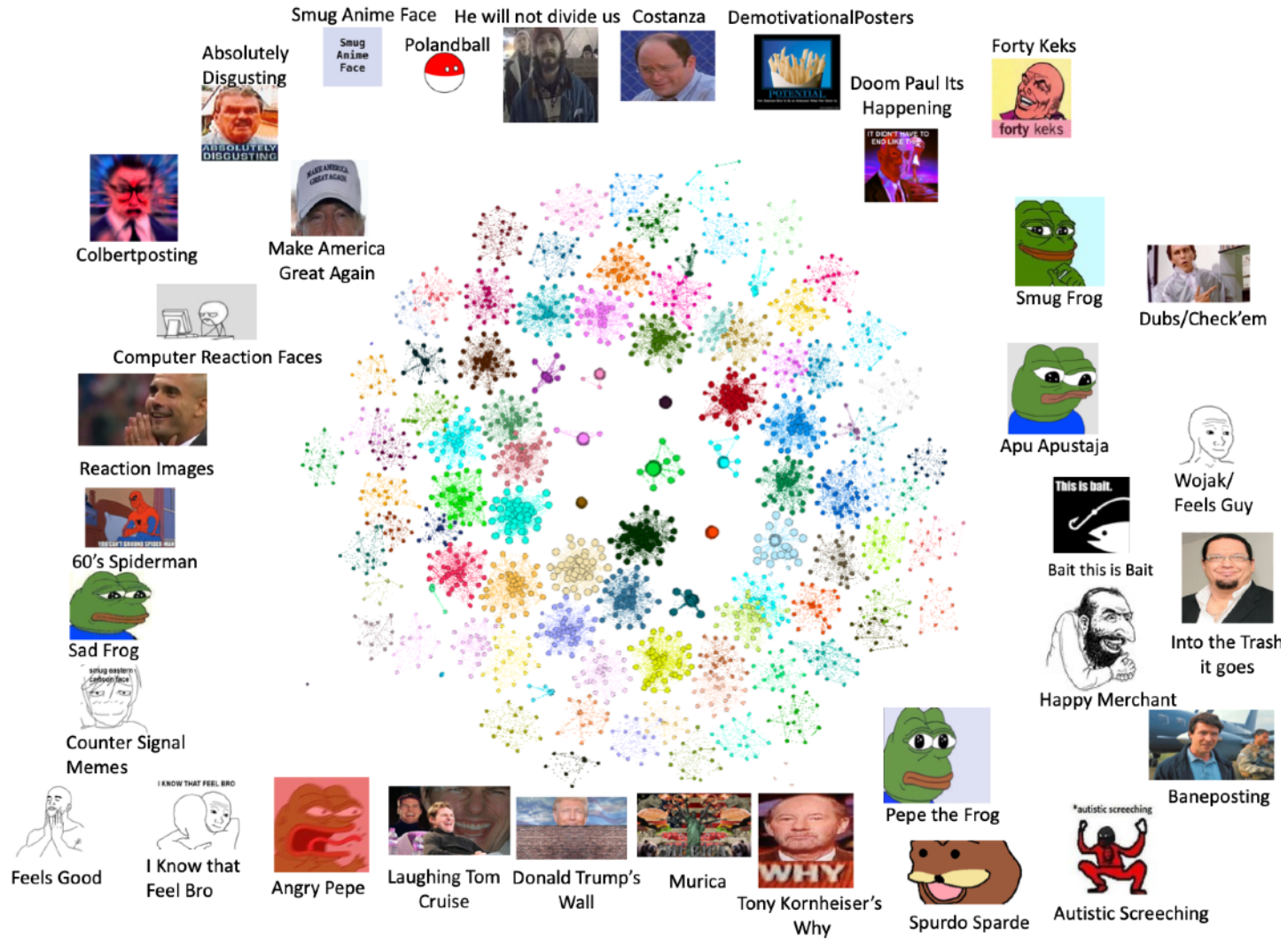


Gab

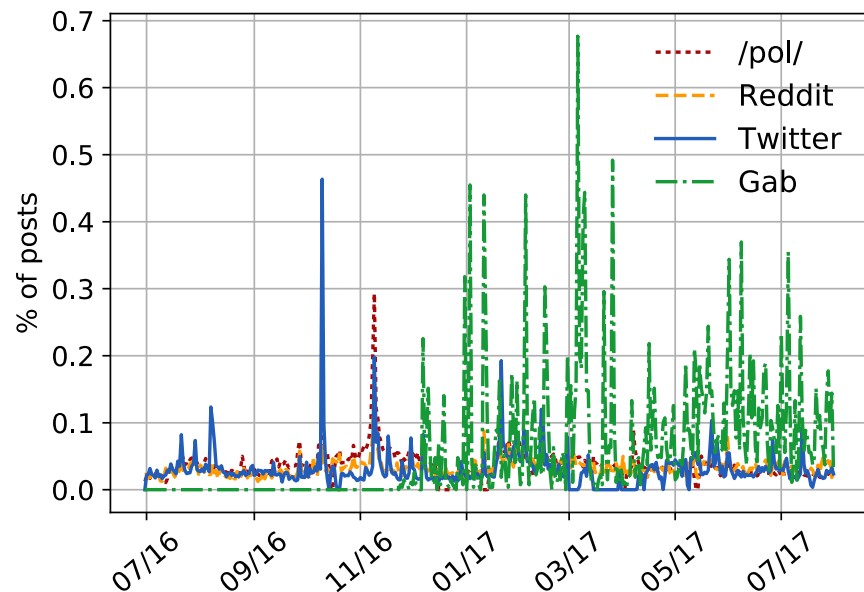


Twitter

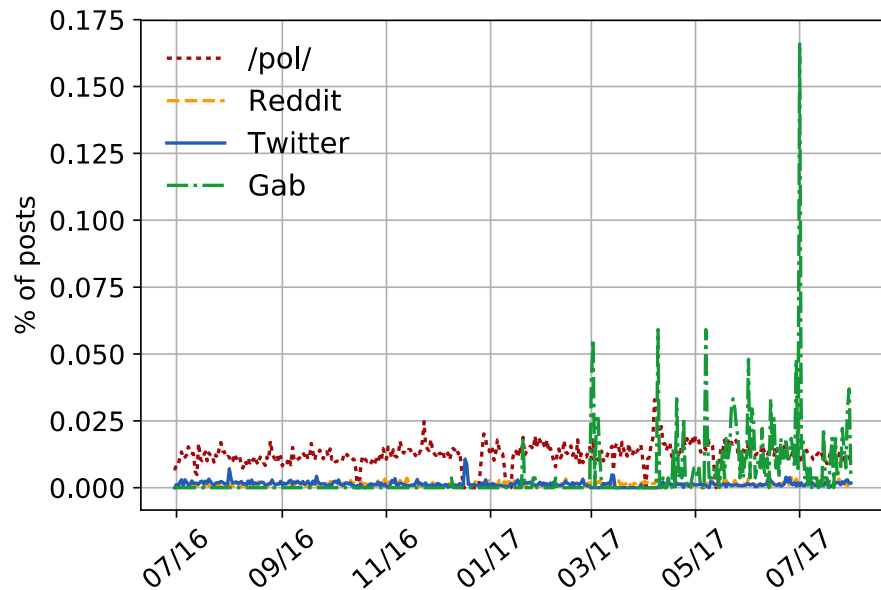




# How are memes shared over time?



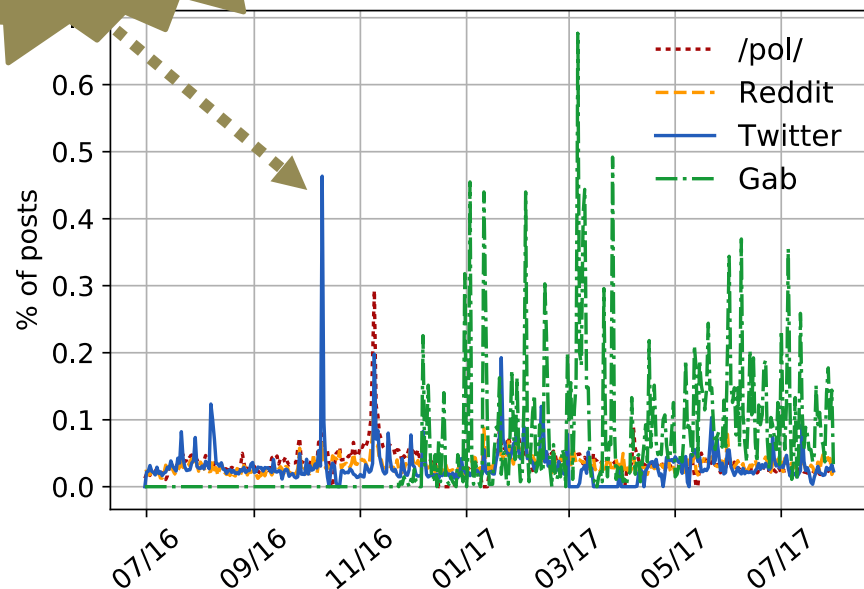
Political Memes



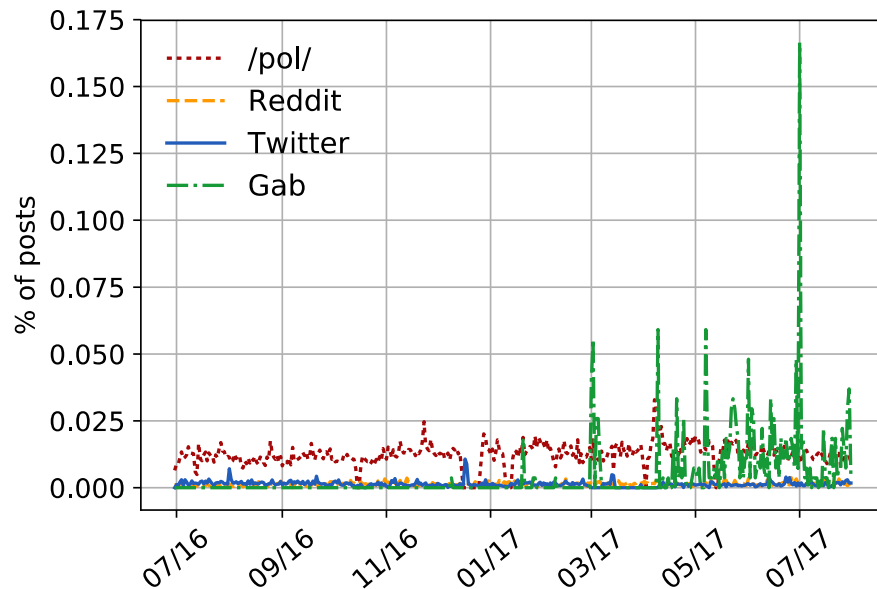
Racist Memes

# How are memes shared over time?

2nd US  
presidential  
debate



Political Memes

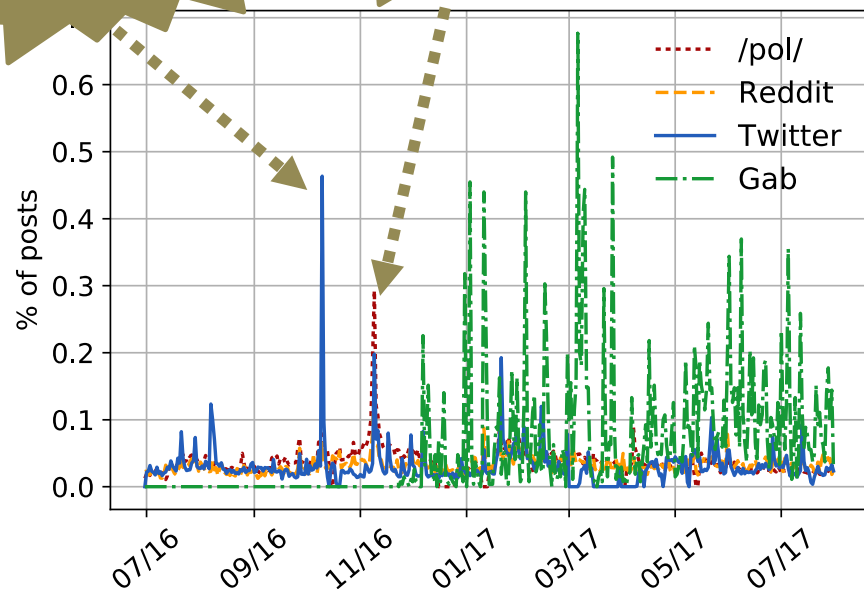


Racist Memes

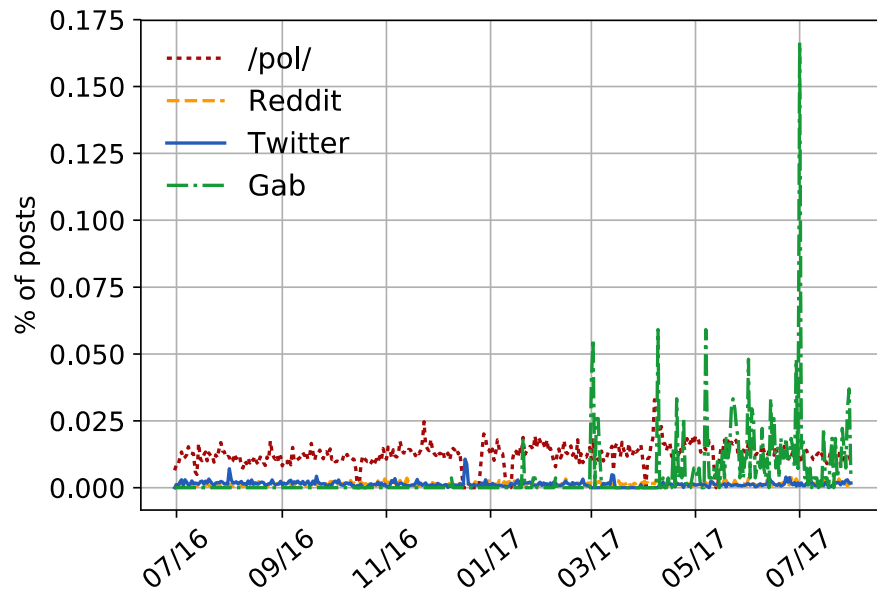
# How are memes shared over time?

2nd US  
presidential  
debate

2016 US  
elections



Political Memes



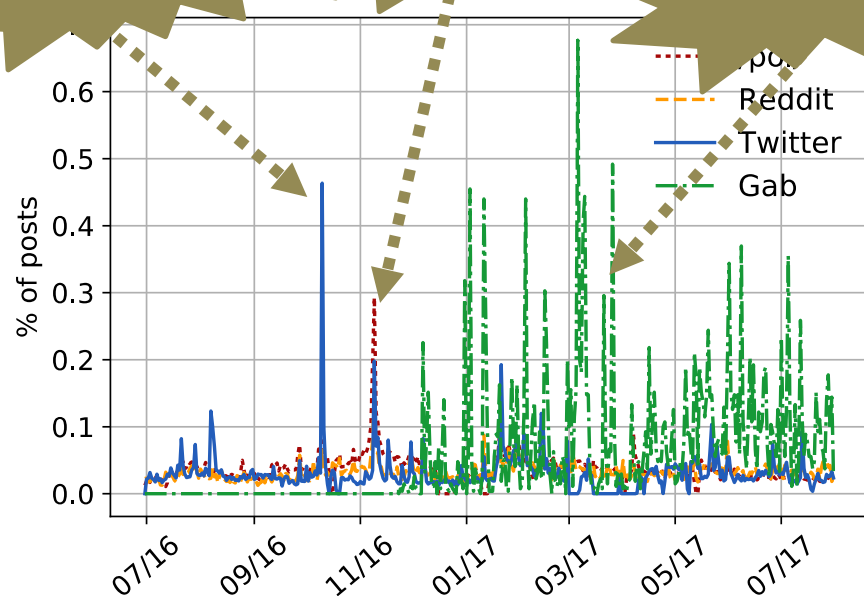
Racist Memes

# How are memes shared over time?

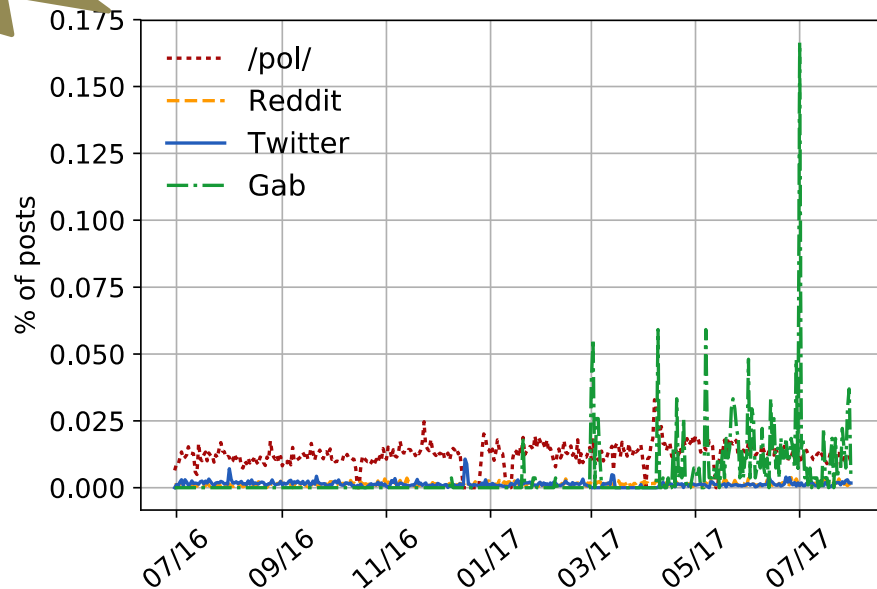
2nd US  
presidential  
debate

2016 US  
elections

Gab activity  
increase  
2017



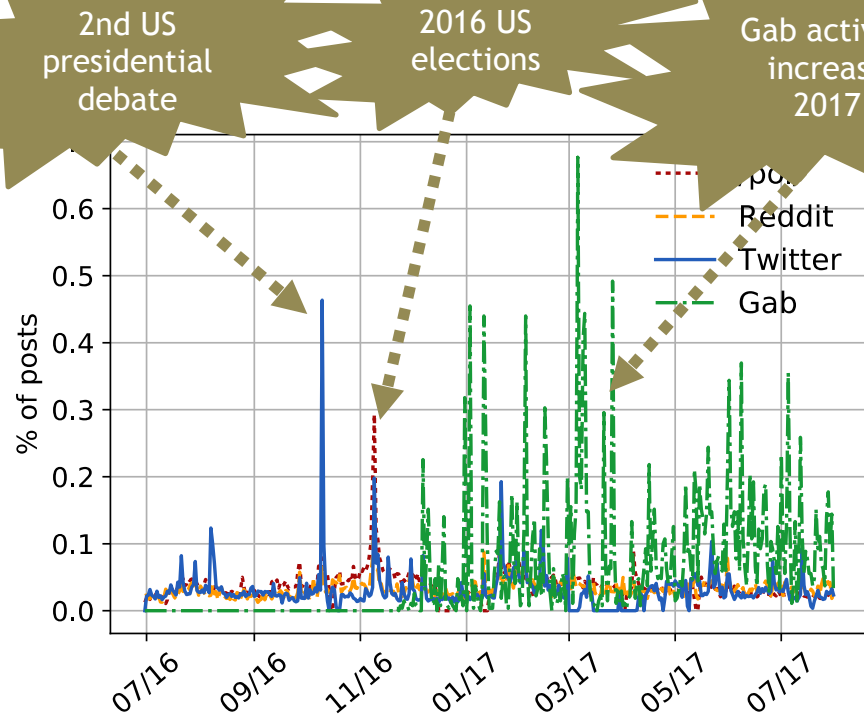
Political Memes



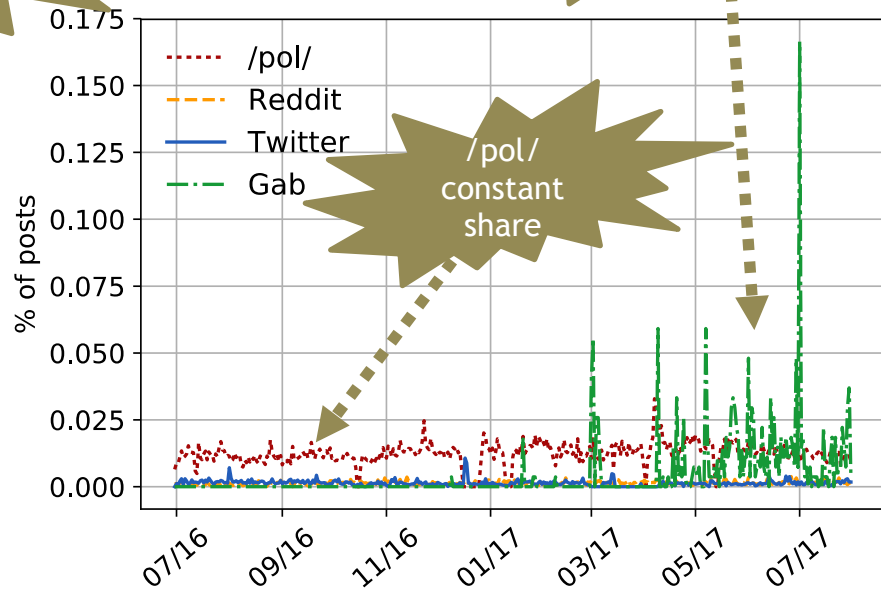
Racist Memes



# 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	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?

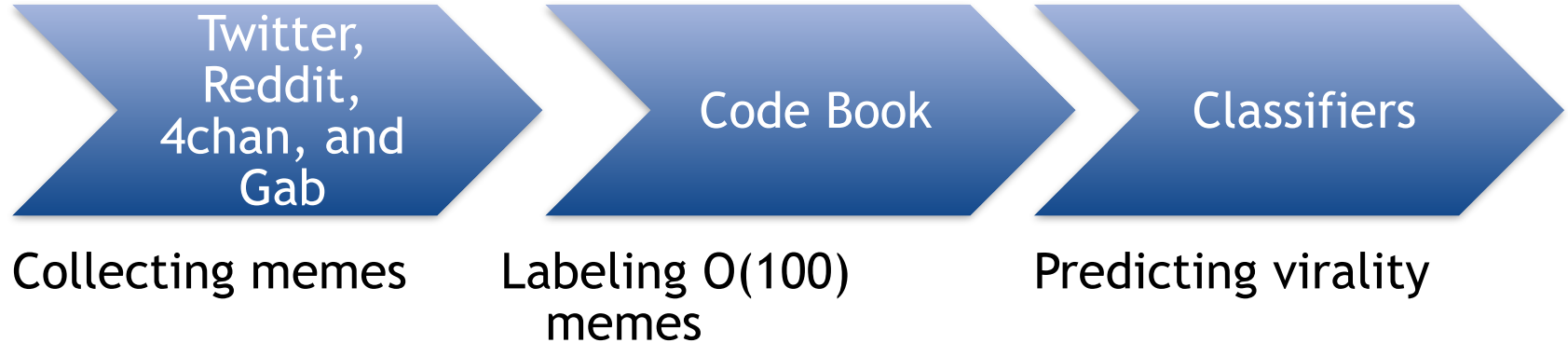
1. Composition
2. Subjects
3. Target audience



i) A viral meme.



ii) A non-viral meme



# TL;DR

- 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>

# 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](https://twitter.com/gianluca_string)

Reginos 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://inews.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](https://twitter.com/jblackb?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor)

>Emiliano De Cristofaro

[https://twitter.com/emilianodc\\_](https://twitter.com/emilianodc_)

>Gianluca Stringhini

[https://twitter.com/gianluca\\_string](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

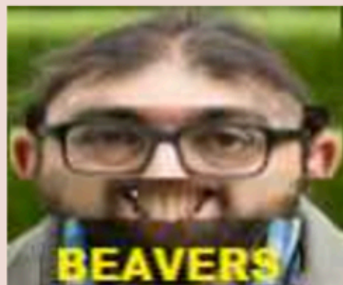
3. South Korean researchers lobby  
government to lift human-embryo  
restrictions

Nature | 08 September 2017

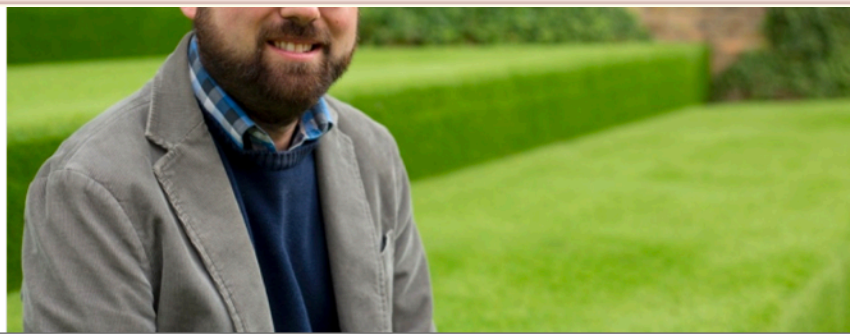


☐ Anonymous (ID: 0a1kdjN4 )  06/10/17(Sat)02:11:32 No.129265175 ► >>129270300

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

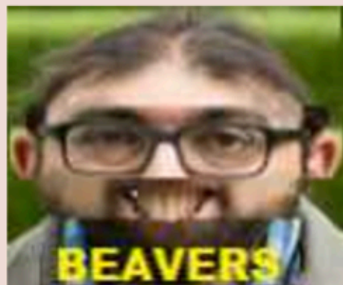


***MOSHE P.I.***

***The Cyberhate  
Investigator***

☐ Anonymous (ID: 0alkdjN4

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



>

nature  
briefing





natur

Home | News & Comm

News & Comment > N

NATURE | NEWS: Q&A

Shining a lig

Cybercrime research

hello i am  
italian



***MOSHE P.I.***

***The Cyberhate  
Investigator***

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





# **Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web\***

Gabriel Emile Hine<sup>‡</sup>, Jeremiah Onaolapo<sup>†</sup>, Emiliano De Cristofaro<sup>†</sup>, Nicolas Kourtellis<sup>#</sup>,  
Ilias Leontiadis<sup>#</sup>, Riginos Samaras<sup>\*</sup>, Gianluca Stringhini<sup>†</sup>, Jeremy Blackburn<sup>#</sup>

# **Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web\***

## **“You Know What to Do”: Proactive Detection of YouTube Videos Targeted by Coordinated Hate Attacks\***

Enrico Mariconti

UCL

e.mariconti@ucl.ac.uk

Guillermo Suarez-Tangil

King's College London

guillermo.suarez-tangil@kcl.ac.uk

Jeremy Blackburn

Binghamton University

blackburn@cs.binghamton.edu

Emiliano De Cristofaro

UCL & Alan Turing Institute

e.decristofaro@ucl.ac.uk

Nicolas Kourtellis

Telefonica Research

nicolas.kourtellis@telefonica.com

Ilias Leontiadis

Samsung AI

i.leontiadis@samsung.com

**Kek, Cucks, and God Emperor Trump: A Measurement Study of  
4chan's Politically Incorrect Forum and Its Effects on the Web\***

**“You Know What to Do”: Proactive Detection of YouTube Videos  
Targeted by Coordinated Hate Attacks\***

**The Web Centipede: Understanding How Web Communities  
Influence Each Other Through the Lens of Mainstream and  
Alternative News Sources**

Savvas Zannettou<sup>★</sup>, Tristan Caulfield<sup>†</sup>, Emiliano De Cristofaro<sup>†</sup>, Nicolas Kourtellis<sup>‡</sup>,  
Ilias Leontiadis<sup>‡</sup>, Michael Sirivianos<sup>★</sup>, Gianluca Stringhini<sup>†</sup>, and Jeremy Blackburn<sup>+</sup>

**Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web\***

**“You Know What to Do”: Proactive Detection of YouTube Videos Targeted by Coordinated Hate Attacks\***

**The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and**

**On the Origins of Memes by Means of Fringe Web Communities\***

Savvas Zannettou\*, Tristan Caulfield<sup>‡</sup>, Jeremy Blackburn<sup>†</sup>, Emiliano De Cristofaro<sup>‡</sup>,  
Michael Sirivianos\*, Gianluca Stringhini<sup>‡◊</sup>, and Guillermo Suarez-Tangil<sup>‡+</sup>

<sup>\*</sup>Cyprus University of Technology, <sup>‡</sup>University College London

<sup>†</sup>University of Alabama at Birmingham, <sup>◊</sup>Boston University, <sup>+</sup>King's College London

# ***“And We Will Fight For Our Race!”* A Measurement Study of Genetic Testing Conversations on Reddit and 4chan\***

Alexandros Mittos<sup>1</sup>, Savvas Zannettou<sup>2</sup>, Jeremy Blackburn<sup>3</sup>, and Emiliano De Cristofaro<sup>1</sup>

<sup>1</sup>University College London   <sup>2</sup>Cyprus University of Technology   <sup>3</sup>Binghamton University

# ***“And We Will Fight For Our Race!”* A Measurement Study of Genetic Testing Conversations on Reddit and 4chan\***

## **Detecting Cyberbullying and Cyberaggression in Social Media\***

Despoina Chatzakou<sup>1</sup>, Ilias Leontiadis<sup>2</sup>, Jeremy Blackburn<sup>3</sup>, Emiliano De Cristofaro<sup>4</sup>,  
Gianluca Stringhini<sup>5</sup>, Athena Vakali<sup>6</sup>, and Nicolas Kourtellis<sup>7</sup>

<sup>1</sup>Center for Research and Technology Hellas, <sup>2</sup>Samsung AI, <sup>3</sup>SUNY Binghamton,  
<sup>4</sup>UCL, <sup>5</sup>Boston University, <sup>6</sup>Aristotle University of Thessaloniki, <sup>7</sup>Telefonica Research

***“And We Will Fight For Our Race!”* A Measurement Study of Genetic Testing Conversations on Reddit and 4chan\***

**Detecting Cyberbullying and Cyberaggression in Social Media\***

**Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web\***

Savvas Zannettou<sup>1</sup>, Tristan Caulfield<sup>2</sup>, Emiliano De Cristofaro<sup>2</sup>,  
Michael Sirivianos<sup>1</sup>, Gianluca Stringhini<sup>3</sup>, Jeremy Blackburn<sup>4</sup>

<sup>1</sup>Cyprus University of Technology, <sup>2</sup>University College London, <sup>3</sup>Boston University, <sup>4</sup>University of Alabama at Birmingham

***“And We Will Fight For Our Race!”* A Measurement Study of Genetic Testing Conversations on Reddit and 4chan\***

**Detecting Cyberbullying and Cyberaggression in Social Media\***

**Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web\***

**Characterizing the Use of Images in State-Sponsored Information Warfare Operations by Russian Trolls on Twitter\***

Savvas Zannettou<sup>1</sup>, Tristan Caulfield<sup>2</sup>, Barry Bradlyn<sup>3</sup>,  
Emiliano De Cristofaro<sup>2</sup>, Gianluca Stringhini<sup>4</sup>, and Jeremy Blackburn<sup>5</sup>





# Understanding the Use of Fauxtography on Social Media

Yuping Wang<sup>♣</sup>, Fatemeh Tahmasbi<sup>♦</sup>, Jeremy Blackburn<sup>♦</sup>, Barry Bradlyn<sup>♣</sup>,  
Emiliano De Cristofaro<sup>♥</sup>, David Magerman<sup>∂</sup>, Savvas Zannettou<sup>★</sup>, and Gianluca Stringhini<sup>♣</sup>

<sup>♣</sup>Boston University, <sup>♦</sup>Binghamton University, <sup>♣</sup>University of Illinois at Urbana–Champaign,  
<sup>♥</sup>University College London, <sup>∂</sup>Differential Venture Partners, <sup>★</sup>Max Planck Institute for Informatics

# Understanding the Use of Fauxtography on Social Media

## The Evolution of the Manosphere Across the Web\*

Manoel Horta Ribeiro<sup>♠</sup>, Jeremy Blackburn<sup>◇</sup>, Barry Bradlyn<sup>◇</sup>, Emiliano De Cristofaro<sup>♣</sup>  
Gianluca Stringhini<sup>♠</sup>, Summer Long<sup>◇</sup>, Stephanie Greenberg<sup>◇</sup>, Savvas Zannettou<sup>◇</sup>

<sup>♠</sup>EPFL, <sup>◇</sup>Binghamton University, <sup>◇</sup>University of Illinois at Urbana-Champaign

<sup>♣</sup>University College London, <sup>♠</sup>Boston University, <sup>◇</sup>Max Planck Institute for Informatics  
– iDRAMA Lab –

manoel.hortaribeiro@epfl.ch, {jblackbu,slong12,sgreenb8}@binghamton.edu,  
bbradlyn@illinois.edu, gian@bu.edu, e.decrisofaro@ucl.ac.uk, szannett@mpi-inf.mpg.de

**Understanding the Use of Fauxtography on Social Media**

**The Evolution of the Manosphere Across the Web\***

**Dissecting the Meme Magic: Understanding Indicators  
of Virality in Image Memes\***

Chen Ling<sup>†</sup>, Ihab AbuHilal<sup>‡</sup>, Jeremy Blackburn<sup>‡</sup>, Emiliano De Cristofaro<sup>⊞</sup>,  
Savvas Zannettou<sup>◇</sup>, and Gianluca Stringhini<sup>†</sup>

<sup>†</sup>Boston University, <sup>‡</sup>Binghamton University, <sup>⊞</sup>University College London, <sup>◇</sup>Max Planck Institute for Informatics  
ccling@bu.edu, iabuhil1@binghamton.edu, jblackbu@binghamton.edu,  
e.decrisofaro@ucl.ac.uk, szannett@mpi-inf.mpg.de, gian@bu.edu

**Understanding the Use of Fauxtography on Social Media**

**The Evolution of the Manosphere Across the Web\***

**Dissecting the Meme Magic: Understanding Indicators  
of Virality in Image Memes\***

**“Is it a Qoincidence?”: A First Step Towards Understanding and  
Characterizing the QAnon Movement on Voat.co**

Antonis Papasavva<sup>1</sup>, Jeremy Blackburn<sup>2</sup> Gianluca Stringhini<sup>3</sup>, Savvas Zannettou<sup>4</sup>,  
and Emiliano De Cristofaro<sup>1</sup>

<sup>1</sup>University College London, <sup>2</sup>Binghamton University, <sup>3</sup>Boston University, <sup>4</sup>Max Planck Institute for Informatics  
antonis.papasavva@ucl.ac.uk, jblackbu@binghamton.edu, gian@bu.edu, szannett@mpi-inf.mpg.de, e.decrisofaro@ucl.ac.uk

– iDRAMA Lab –

# Thank you!

