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

Emiliano De Cristofaro
genic@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, ...)

mitigating unwanted activity
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

Over the past 24 hours, Leslie Jones has been inundated with racist, hateful vitriol on Twitter.
How a racist, silver-tongued mob forced Leslie Jones off Twitter

By Kristen V. Brown

SONY

INTERNET
8 DECEMBER 2016

Pizzagate: How a 4Chan conspiracy went mainstream

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

Over the past 24 hours, Leslie Jones has been inundated with racist, hateful vitriol on Twitter.
How a racist, sick mob forced Leslie Jones off Twitter

By Kristin V. Brown

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

Over the past 24 hours, Leslie Jones has been inundated with racist, hateful vitriol on Twitter.
Anti-Muslim Hate Has Been Rampant on Reddit Since the New Zealand Shooting

And the platform doesn't seem to care.

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

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.
Heard of 4chan?
Heard of 4chan?
Heard of 4chan?

I CAN HAS CHEEZBURGER?
Heard of 4chan?
Heard of 4chan?
Heard of 4chan?

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 (70 at the moment)

An “original post” (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.

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

Anonymous (ID: lv9eTFJC) 01/08/15(Thu)23:40:44 No.40489590

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 Board

Administrators (ID: lv9, lv6) at 01/01/2023, 23:40:44 No. 40489590

This thread is for general discussions, 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. While 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

![Graph showing CDF of TOXICITY and SEVERE_TOXICITY](image1)

![Graph showing CDF of INFLAMMATORY, PROFANITY, INSULT, and OBSCENE](image2)
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 platforms (e.g., YouTube)

Hans2k4ever 1 month ago
When will the normies learn?
Reply • 181

View all 3 replies ▼

Jack Harper 1 month ago
mfw he accidentally rephills 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
Activity Peaks

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

YT videos with peaks during 4chan thread

Determined via PDF of commenting timeseries
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

![Graph showing synchronization lag against hate comments per second and overlap in commenters.]

- YT comments have hate
- YT comments have no hate
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%
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
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

WikiLeaks

Data Provider

4chan

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

WikiLeads

Data Provider

4chan

Theory Provider

Reddit

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
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
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 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%)

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

The vehicle seized by law enforcement in Florida was covered in strange propaganda.
Memes have become a popular, and seemingly effective, method to transmit ideology.

Memes have been weaponized
Memes processing pipeline

Data Collection (160M images) → Perceptual Hashing → Clustering → Influence Estimation (Hawkes)

Code for our analysis pipeline (and data) available at https://github.com/memespaper/memes_pipeline
Examples of clusters
# Top memes per social network

<table>
<thead>
<tr>
<th>4chan (/pol/)</th>
<th>Reddit</th>
<th>Gab</th>
<th>Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
</tbody>
</table>
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
  - 2nd US presidential debate
  - 2016 US elections
  - Gab activity increase 2017

- Racist Memes
  - Gab activity increase in 2017
  - /pol/ constant share

Graphs showing the percentage of posts for political and racist memes over time from 07/16 to 07/17, with peaks during specific events.
Communities’ influence (racist memes)

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

<table>
<thead>
<tr>
<th>Source</th>
<th>/​pol/</th>
<th>Reddit</th>
<th>Twitter</th>
<th>Gab</th>
<th>T_D</th>
</tr>
</thead>
<tbody>
<tr>
<td>R: 99.44%</td>
<td>R: 6.89%</td>
<td>R: 3.94%</td>
<td>R: 18.93%</td>
<td>R: 14.17%</td>
<td></td>
</tr>
<tr>
<td>NR: 97.12%*</td>
<td>NR: 3.95%*</td>
<td>NR: 2.93%*</td>
<td>NR: 12.91%</td>
<td>NR: 16.41%*</td>
<td></td>
</tr>
<tr>
<td>R: 0.29%</td>
<td>R: 88.95%</td>
<td>R: 2.79%</td>
<td>R: 1.25%</td>
<td>R: 9.84%</td>
<td></td>
</tr>
<tr>
<td>NR: 1.28%*</td>
<td>NR: 90.87%*</td>
<td>NR: 4.64%*</td>
<td>NR: 9.42%</td>
<td>NR: 9.02%*</td>
<td></td>
</tr>
<tr>
<td>R: 0.15%</td>
<td>R: 1.75%</td>
<td>R: 92.95%</td>
<td>R: 0.66%</td>
<td>R: 1.37%</td>
<td></td>
</tr>
<tr>
<td>NR: 0.79%</td>
<td>NR: 2.85%*</td>
<td>NR: 90.96%*</td>
<td>NR: 8.16%</td>
<td>NR: 3.66%*</td>
<td></td>
</tr>
<tr>
<td>R: 0.04%</td>
<td>R: 0.65%</td>
<td>R: 0.10%</td>
<td>R: 76.67%</td>
<td>R: 0.08%</td>
<td></td>
</tr>
<tr>
<td>NR: 0.09%</td>
<td>NR: 0.16%</td>
<td>NR: 0.20%</td>
<td>NR: 59.72%</td>
<td>NR: 0.58%</td>
<td></td>
</tr>
<tr>
<td>R: 0.07%</td>
<td>R: 1.75%</td>
<td>R: 0.23%</td>
<td>R: 2.49%</td>
<td>R: 74.53%</td>
<td></td>
</tr>
<tr>
<td>NR: 0.73%*</td>
<td>NR: 2.18%*</td>
<td>NR: 1.27%*</td>
<td>NR: 9.79%</td>
<td>NR: 70.32%*</td>
<td></td>
</tr>
</tbody>
</table>
If we look at the influence normalized to the number of memes posted, the The_Donald is most efficient in terms of disseminating memes.
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
Methodology

Twitter, Reddit, 4chan, and Gab

Collecting memes

Code Book

Labeling O(100) memes

Classifiers

Predicting virality

i) A viral meme. ii) A non-viral meme
• 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

What about state-level adversaries?
What about state-level adversaries?
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?

Nudging users to follow them

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

News accounts? Trump supporters
Where are they allegedly posting from?
What hashtags they shared?

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

One month before 2016 US elections
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%.
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
/ˈma-nə-sfɜə/

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

(Lily 2016)
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.
Idea

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

52 subreddits
6 standalone forums
835k users
22M posts

Available @ zenodo.org/record/4007913
RQ1 How has the popularity and the levels of activity of the different Manosphere communities evolved?
Activity on Reddit

%Posts

Year

Activity on Reddit

Men's Rights Activists

Pick Up Artists

"Red Pill"

Incels

MGTOV
Activity on Reddit

Men's Rights Activists
Pick Up Artists
"Red Pill"
Incels
MGTOW

%Posts

Year

0%
25%
50%
75%
100%

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?
MRA @ Year i

X

MG TOW @ Year i+1
Migration Relative to Source = \[ \frac{X}{\text{MRA @ Year } i} \]

\[ \text{MRA @ Year } i \cap \text{MGTOW @ Year } i+1 \]
Migration Relative to Source = \( \frac{X}{\text{MRA @ Year } i} \)

Migration Relative to Destination = \( \frac{X}{\text{MGTOW @ Year } i + 1} \)

\[ \text{MRA @ Year } i \quad \text{X} \quad \text{MGTOW @ Year } i + 1 \]
Relative to Destination (MGTOW)
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...
Takeaway #3

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

> his work on the contagious aspects of cheating and understanding and learning about cyberbullying/hate behavior in online games represents 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

Riggins 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

> 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.”


Jeremiah Onanalo

> Just a henchman but...

> Racist or forthcoming papers: "What Happens After You Are Pwned: 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://www.UCL.ac.uk/staff/J.Onanalo/papers/honeymail.pdf

Gabriel Emile Hime

> no joke

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/Jblackburn?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor

> Emiliano De Cristofaro

https://twitter.com/emilianodic...

> 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
Shining a light on the dark corners of the web

Cybercrime researcher Gianluca Stringhini explains how he studies bots, speech and fake news.
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
Provided inspiration for new memes

Figure 29: A very comfy Savvas Zanettou
Thank you!