Studying jerks on the Internet: A Data-Driven Approach

Emiliano De Cristofaro

(Thanks to Jeremy Blackburn and Savvas Zannettou for most of the slides)
Outline

1. Hate on and “raids” from fringe communities like 4chan [ICWSM 2017]

2. Influence of fringe communities on misinformation [IMC 2017]

3. Misuse of Web archiving services [ICWSM 2018]

More work on online hate, cyberbullying, etc: https://encase.socialcomputing.eu/publications
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?

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

Anonymous (ID: [REDACTED])
09/11/16(Sun)22:27:40 No.88456696

HILLAREEEEE

115 KB JPG

>>88456696 #

Anonymous (ID: [REDACTED])
09/11/16(Sun)22:31:21 No.88456696

>>88456217 #
Get this cuck outta here. He isn’t a man of conviction and virility like Cruz. He’s just a sad little man who slobbers the spit of Hillary’s shoes. I love Trump but I respect people who have true beliefs and a moral compass and stand up for themselves.

>>88457099 #

Anonymous (ID: [REDACTED])
09/12/16(Mon)03:24:29 No.891806

Clinton Gone = Trump’s Worst N(...) 09/12/16(Mon)03:24:29 No.891806

You faggots realize that Clinton being ill/dying isn’t a good thing for Trump, right?

If she decides to drop out of the race then what? The Dems put up Kaine, Biden, or maybe even Sanders.

Then Trump gets BTFO in a landslide loss because the Dem’s formerly utterly hated and untrustworthy candidate is gone and replaced by someone much more likable.

It’s as if you’re all actually shortsighted and retarded and you’re not all pretending.

433 KB PNG

53 KB PNG

>>88457099 #

Anonymous (ID: [REDACTED])
09/12/16(Mon)07:02:57 No.88514002

>>88512762 (OP) #
Gas the Kikes, race war now.

1.41 MB PNG

>>88514099 # >>88514299 #

Anonymous (ID: [REDACTED])
09/12/16(Mon)07:03:58 No.88514099

>>88514002 #
First post tits post.

Anonymous (ID: [REDACTED])
09/09/16(Fri)07:25:20 No.815167

This is a holy thread. Please pay respects to Kek and your post will be blessed.

64 KB JPG
Heard of 4chan?

I CAN HAS CHEEZBURGER?

WHERE DID EVERYTHING GO SO REICH
Why Do We Care About 4Chan?

“4chan is niche and irrelevant” – Anonymous WWW’17 Reviewer
How a racist, sexist hate mob forced Leslie Jones off Twitter

Leslie Jones hack: Inside the alt-right's racist celebration of Jones' humiliation

4chan may have remotely wiped a Clinton official's iPhone and iPad

#MediasNextTrumpLie: Inside 4chan's pro-Trump spam campaign to derail the media

Who turned Microsoft's chatbot racist? Surprise, it was 4chan and 8chan
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

**Extremely** lax moderation

Volunteer “janitors” as well as ”admins”

*Almost anything goes*
In This Talk: Challenges of Measuring 4chan

1. Not your typical social network (anonymous/ephemeral)
2. Their actions not limited to 4chan, need to look at other platforms to measure their impact
3. Knowing what they’re talking about is not easy
4. You get raided
5. You might get “redpilled”
Anonymity & Ephemerality

Users do not need to register an account to participate
   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
   (More later)
Datasets

<table>
<thead>
<tr>
<th></th>
<th>/pol/</th>
<th>/sp/</th>
<th>/int/</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threads</td>
<td>217K</td>
<td>14.4K</td>
<td>24.9K</td>
<td>256K</td>
</tr>
<tr>
<td>Posts</td>
<td>8.3M</td>
<td>1.2M</td>
<td>1.4M</td>
<td>10.9M</td>
</tr>
</tbody>
</table>

June 30 to September 12, 2016

Methodology:
Visit the “catalog”
Take a snapshot every 5 minutes
Once a thread is pruned, retrieve full/final contents from archive

We’re still crawling…
Ephemerality: The Bump System

Limit boards to N live threads

Threads ordered by MRU

A new post in a thread “bumps” it up to the top

Create new thread $\rightarrow$ Old thread dies

Bump limit

Max times thread can be bumped

No discussion will dominate forever
/pol/ users seems well distributed

Native English speaking countries most highly represented
Plenty of other countries really well represented too though!
Are Flags Trustworthy?

Use spectral clustering of the topics that each country posts about.

The clusters follow real world socio-political blocks.

While flags are not perfect, they seem reasonable.
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%

An Anonymous user (ID: StmUFWN) commented:

Can't believe I sat and read all this shit, most interesting part was the countries that get the most replies.

Trolls now have the list of the top 10 countries to use to bait.

Also, the fact "trash" is the 10th most popular "hate speech" word, fucking burgers and leaves.
Raids

Attempts to disrupt another site

*Not* a DDoS

Disrupts *community* that calls service home, not the service itself

Raids are a favorite past time on 4chan

“Pool’s closed!”

Have become less “funny” and more “scary” lately

It’s a *socio-technical* problem
YouTube Raids

Someone posts a YouTube link
  Maybe with a prompt like “you know what to do”

Thread is an aggregation point for raiders
  E.g., “Hah! I called that person a nigger!”

If raid is taking place:
  Peak in YouTube comments while thread alive?
  /pol/ thread and YT comments *synchronized*?
Raids

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

Emiliano De Cristofaro
11,013 views

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

Blue lines $\rightarrow$ per-sample lag
Red area $\rightarrow$ density of the lags
Peak of density curve = 0.2s
Validation

YT comments have hate
YT comments have no hate

Anonymous (ID: DaWZlbh)  10/13/16(Thu)20:08:55 No.92772435

>>92765569
Physicist finalist here, can confirm the maths is legitimate. Holy fuck
When U.N. sends its Pepes, they’re not sending their best. They’re not sending Kek, they’re not sending rares. They’re sending Pepes that have lots of copies, and they’re bringing those copies with us. They’re bringing tumblr. They’re bringing facebook. They’re common. And some, I assume, are rare Pepes.
The Cuckedemics, A Study Anonymous (ID: zw3wNfrR) 10/13/16(Thu)06:19:21 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 not just for the gaming community, but in understanding anti-social behavior in the real-world as well.
http://www.tdi.es/research/researchers/J-blackburn

Gianluca Stringhini
> Twitter Cuck e.g. “But #brexit has nothing to do with xenophobia, right?”
> Retweets Hilary
https://twitter.com/gianluca_string

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

Ilias Leonidakis
> bizarre that they would include this guy because his papers are all related to internet privacy and ad reverse.
> 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.tdi.es/research/researchers/Ilias-Leonidakis

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.

Jeremiah Onaloapo
> Just a henchman but...
> Recent 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.w0.lcs.ucl.ac.uk/staff/J.Onaloapo/papers/imc-15gmail.pdf

Gabriel Emilie 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/jblackbo
> ref src tw src %SEgoogle%CTwcred%SEsep%CTwgr%SEauth

Emiliano De Cristofaro
https://twitter.com/emiliandecr

Gianluca Stringhini
https://twitter.com/gianluca_string
Outline

1. Hate on and “raids” from fringe communities like 4chan
   [ICWSM 2017]

2. Influence of fringe communities on misinformation
   [IMC 2017]

3. Misuse of Web archiving services
   [ICWSM 2018]
Can We Quantify How Rest of Web is Affected?

4chan

Reddit

Twitter

Facebook
The Information Ecosystem
4chan → Twitter

We should start an immediate investigation into @SenSchumer and his ties to Russia and Putin. A total hypocrite!

Here is picture of Senator Chuck Schumer having coffee and donuts with Vladimir Putin.

We need to investigate this meeting and any Russian ties with Chuck.
Donald J. Trump
@realDonaldTrump

#FraudNewsCNN #FNN

4:21 PM - Jul 2, 2017

141,936 368,969 605,362

Vox

Trump's anti-CNN tweet originated from Reddit’s largest right-wing extremist forum

Regardless of his intentions, Trump's tweet aligned him with violent hate speech.

Updated by Aja Romano | @ajaromano | Jul 3, 2017, 5:00pm EDT
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
Oh Hey... It’s 4chan Again...

Fake news of a fatal car crash wiped out $4 billion in Ethereum’s market value yesterday

Geary Danley was not the gunman in Las Vegas who killed at least 50 people late Sunday. But for hours on the far-right internet, would-be sleuths scoured Danley’s Facebook likes, family photographs and marin...
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
The Data

99 mainstream and alternative ("fake") news sources

<table>
<thead>
<tr>
<th>Platform</th>
<th>Posts/Comments</th>
<th>Alternative URLs</th>
<th>Mainstream URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>486K</td>
<td>42K</td>
<td>236K</td>
</tr>
<tr>
<td>Reddit (six selected subreddits)</td>
<td>620K</td>
<td>40K</td>
<td>301K</td>
</tr>
<tr>
<td>4chan (/pol/)</td>
<td>90K</td>
<td>9K</td>
<td>40K</td>
</tr>
</tbody>
</table>
Here’s What The Graph Looks Like

Mainstream News Sources

Alternative News Sources
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 in other processes
   Increases the rates of other processes for a period of time

Enables us to be confident about the number of events caused by another event on the source process (weight)
   Reveals causal relationships
Hawkes Processes Example
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
What 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%)
What Communities Influence Each Other?

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

1. Hate on and “raids” from fringe communities like 4chan [ICWSM 2017]
2. Influence of fringe communities on misinformation [IMC 2017]
3. Misuse of Web archiving services [ICWSM 2018]
Web archiving

Archive.is

On-demand archiving service
Generates 5-character archive URLs
E.g., www.google.de becomes archive.is/HVbU

Wayback Machine

Works mainly through a proactive crawler
Supports also on-demand archival
E.g., www.google.com becomes
How do Web archiving services work?

A. Archival

1. Source URL
2. Obtain URL’s Source
3. Store frozen copy
4. Archive URL

B. Retrieval

1. Archive URL
2. Frozen copy

Source
How do Web archiving services work?

1. Source URL
2. Obtain URL’s Source
3. Store frozen copy
4. Archive URL

(Possibly) Source URL obfuscation

Content Persistence (Source not involved in retrieval)

“Right to be forgotten”?
Research Questions

1. What kind of content get archived, by whom, and why?

2. How are archive URLs disseminated on the Web?

3. Are archiving services misused?
# Datasets

1. **Live Feed archive.is dataset**

2. **Archive.is and Wayback Machine URLs posted on Reddit, /pol/, Gab, Twitter**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Archive URLs</th>
<th>Source URLs</th>
<th>Source Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Live Feed</strong></td>
<td>Archive.is</td>
<td>21M</td>
<td>20.6M</td>
</tr>
<tr>
<td><strong>Reddit</strong></td>
<td>Archive.is</td>
<td>310K</td>
<td>291K</td>
</tr>
<tr>
<td></td>
<td>Wayback</td>
<td>387K</td>
<td>343K</td>
</tr>
<tr>
<td><strong>/pol/</strong></td>
<td>Archive.is</td>
<td>36K</td>
<td>33K</td>
</tr>
<tr>
<td></td>
<td>Wayback</td>
<td>2K</td>
<td>2K</td>
</tr>
<tr>
<td><strong>Gab</strong></td>
<td>Archive.is</td>
<td>5.9K</td>
<td>5.7K</td>
</tr>
<tr>
<td></td>
<td>Wayback</td>
<td>0.3K</td>
<td>0.3K</td>
</tr>
<tr>
<td><strong>Twitter</strong></td>
<td>Archive.is</td>
<td>3.7K</td>
<td>3.6K</td>
</tr>
<tr>
<td></td>
<td>Wayback</td>
<td>1.2K</td>
<td>1.2K</td>
</tr>
</tbody>
</table>
What kind of content gets archived?

Get categories for domains using Virus Total API

Live feed dataset has 5.3M domains
Select top 100K domains for categorization
What kind of content gets archived? (OSNs)

OSN posts among the top two categories in all Web Communities

News among the top five categories in all Web Communities
Platform-specific censorship

Some domains exclusively shared via archiving services. Why?

8chan and Facebook considered spam (rejected) on /pol/

Users utilize archive.is to circumvent this platform-specific censorship

Evidence of accidental news source censorship because of substitution filters (used for fun)

   smh becomes baka (smh.com.au becomes baka.com.au)
### Original content availability

<table>
<thead>
<tr>
<th>platform</th>
<th>% of unavailable URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reddit</td>
<td></td>
</tr>
<tr>
<td>Archive.is</td>
<td>7%</td>
</tr>
<tr>
<td>Wayback</td>
<td>11%</td>
</tr>
<tr>
<td>/pol/</td>
<td></td>
</tr>
<tr>
<td>Archive.is</td>
<td>18%</td>
</tr>
<tr>
<td>Wayback</td>
<td>34%</td>
</tr>
<tr>
<td>Gab</td>
<td></td>
</tr>
<tr>
<td>Archive.is</td>
<td>13%</td>
</tr>
<tr>
<td>Wayback</td>
<td>52%</td>
</tr>
<tr>
<td>Twitter</td>
<td></td>
</tr>
<tr>
<td>Archive.is</td>
<td>24%</td>
</tr>
<tr>
<td>Wayback</td>
<td>51%</td>
</tr>
</tbody>
</table>
Who is archiving URLs? Reddit

31% and 82% of archive.is and Wayback Machine URLs, resp., are posted by “SnapshillBot”

44% of archive.is URLs are shared by bots, 85% for Wayback Machine

→ Content persistence within subreddits?
Speaking of subreddits.....

<table>
<thead>
<tr>
<th>Subreddit (archive.is)</th>
<th>(%)</th>
<th>Subreddit (Wayback)</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The_Donald</td>
<td>24.48%</td>
<td>EnoughTrumpSpam</td>
<td>31.82%</td>
</tr>
<tr>
<td>KotakuInAction</td>
<td>15.83%</td>
<td>MGTOW</td>
<td>7.38%</td>
</tr>
<tr>
<td>EnoughTrumpSpam</td>
<td>12.06%</td>
<td>SnapshillBotEx</td>
<td>7.19%</td>
</tr>
<tr>
<td>MGTOW</td>
<td>3.48%</td>
<td>undelete</td>
<td>5.90%</td>
</tr>
<tr>
<td>undelete</td>
<td>2.74%</td>
<td>SubredditDrama</td>
<td>5.50%</td>
</tr>
<tr>
<td>SubredditDrama</td>
<td>2.61%</td>
<td>Drama</td>
<td>5.03%</td>
</tr>
<tr>
<td>Drama</td>
<td>2.33%</td>
<td>Gamingcirclejerk</td>
<td>3.47%</td>
</tr>
<tr>
<td>Gamingcirclejerk</td>
<td>1.57%</td>
<td>ShitAmericansSay</td>
<td>1.63%</td>
</tr>
<tr>
<td>conspiracy</td>
<td>1.44%</td>
<td>TopMindsOfReddit</td>
<td>1.51%</td>
</tr>
<tr>
<td>MensRights</td>
<td>1.12%</td>
<td>TheBluePill</td>
<td>1.25%</td>
</tr>
<tr>
<td>savedyouaclick</td>
<td>1.00%</td>
<td>Buttcoin_1000</td>
<td>1.15%</td>
</tr>
<tr>
<td>politics</td>
<td>0.98%</td>
<td>AgainstHateSubreddits</td>
<td>1.06%</td>
</tr>
<tr>
<td>DerekSmart</td>
<td>0.76%</td>
<td>subredditcancer</td>
<td>0.99%</td>
</tr>
<tr>
<td>ShitAmericansSay</td>
<td>0.75%</td>
<td>The_Donald</td>
<td>0.95%</td>
</tr>
<tr>
<td>PoliticsAll</td>
<td>0.72%</td>
<td>badeconomics</td>
<td>0.75%</td>
</tr>
</tbody>
</table>
Ad-Revenue Deprivation

Your submission was removed because it is from washingtonpost.com, which has been identified as a severely anti-Trump domain. Please submit a cached link or screenshot when submitting content from this domain. We recommend using http://www.archive.is for this purpose.

I am a bot, and this action was performed automatically. Please contact the moderators of this subreddit if you have any questions or concerns.
News “Censorship”

13K submissions removed from The_Donald subreddit

In total 23 “anti-Trump” news sources censored

<table>
<thead>
<tr>
<th>News Source</th>
<th># of submissions removed</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>washightonpost.com</td>
<td>3.8K</td>
<td>44%</td>
</tr>
<tr>
<td>cnn.com</td>
<td>3.3K</td>
<td>40%</td>
</tr>
<tr>
<td>nydailynews.com</td>
<td>1.0K</td>
<td>46%</td>
</tr>
<tr>
<td>huffingtonpost.com</td>
<td>0.9K</td>
<td>43%</td>
</tr>
<tr>
<td>nationalreview.com</td>
<td>0.7K</td>
<td>45%</td>
</tr>
</tbody>
</table>
Thank you!
CUCKADEMICS

Primary members

Jeremy Blackburn
University of Alabama at Birmingham

Emiliano De Cristofaro
University College London

Guillermo Suarez-Tangil
King’s College London

Nicolas Kourtellis
Telefonica Research

Ilias Leontiadis
Telefonica Research

Michael Sirivianos
Cyprus University of Technology

Gianluca Stringhini
University College London

DRAMANAUTS

Students and other affiliates

Tristan Caulfield
University College London

Despoina Chatzakou
Aristotle University of

Antigoni Founta
Aristotle University of

Gabriel Hine
Roma3 University

Jeremiah Onaolapo
University College London

Savvas Zannettou
Cyprus University of