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

Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, Jeremy Blackburn
Information ecosystem
Motivation
4chan → Twitter

Anonymous

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

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

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

By BBC Trending
What's popular and why

© 2 December 2016

WikiLeaks → 4chan → Theory Incubators & Gateway to mainstream “world” → Large-scale Disseminator
4chan
Background
4chan basics

• Anonymous conversations grouped into threads

• Original Poster (OP) creates a new thread by making a post with an image

• Other users can reply with or without images

• No likes, shares, favorites, etc.
4chan boards and moderation

• Threads are separated into different areas of interests know as **boards**
  o Areas range from politics to sports
  o Extremely lax moderation by volunteers

• We focus on the Politically Incorrect board (/pol/)
Why do we care about 4chan?

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

Geary Danley was named as the suspect in the Las Vegas shooting, but people were quick to point out that he is not the right person. Geary Danley is a far-right troll and not the person who committed the crime.

Vitalik Buterin confirmed dead. Insiders unloading ETH. The notorious online message board frequented by internet trolls. 'Fatal car crash,' it went on. 'Now we have our answer. He was the glue.'
Reddit Background
Reddit basics

• Popular news aggregator
  ○ “Front page of the Internet”

• A user can start a new thread by creating a submission with a URL

• Other users can reply in a structured way with or without URLs

• Users can upvote/downvote submissions and replies
Subreddits

• Thousands of user-created subreddits
  o Interests range from video games to news, and pornography
  o Each subreddit has its own moderation policy

• We focus on 6 subreddits
  o The_Donald, conspiracy, news, worldnews, politics, and AskReddit
Why do we care about Reddit?

Moderators of pro-Trump Reddit group linked to fake news crackdown on posts

- Posts about two conspiracy theories removed from r/the_donald
- Trolls trying ‘to slander and attack our community’

One of the web’s most virulent incubators of fake news stories about the US election has begun a crackdown on conspiracy theories and “outlandish or unverified claims”.

Moderators on a Reddit group r/the_donald, which became a focal point for supporters of Donald Trump during the election, explained on Tuesday morning that they had been removing some posts about two conspiracy theories because they were “drowning out thoughtful discussion or Trump related content”.
Datasets and Analysis
Datasets

- Compiled a list of 99 mainstream and alternative 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>
Temporal analysis

• Studied the appearance of alternative and mainstream URLs within the platforms

• Built a sequence of appearance for each URL according to the timestamps

• Built a graph with the sequences
Graph representation of the news ecosystem

Twitter

/bpol/

6 subreddits

6 subreddits
Hawkes processes

• Consists of K processes
  ○ Each with a rate of events (i.e., posting of a URL), called background rate

• An event can cause impulse responses to other processes
  ○ Increases the rates of other processes for a period of time

• Enable us to be confident about the number of events caused by another event on the source process (weight)
  ○ Reveal causal relationships
Hawkes processes example

Reddit

Twitter

/pol/
Hawkes processes for influence estimation

• Hawkes model with 8 processes
  o One for each platform
  o 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
Influence Estimation Findings

- Twitter top influencers for alternative URLs
  - The_Donald (2.72%)
  - /pol/ (1.96%)
  - Politics (1.1%)

- Twitter top influencers for mainstream URLs
  - Politics (4.29%)
  - /pol/ (3.01%)
  - The_Donald (2.97%)
Conclusions & Future Work

**Analyzed how news propagate across Web communities**
- Considered URLs from 99 mainstream and alternative news sources

**Provided quantifiable influence between**
- Six subreddits within Reddit
- Twitter
- Politically Incorrect (/pol/) board of 4chan

**Future Work**
- Investigate the use of NLP and Image Recognition to associate events that appear in multiple modalities
Thank you!

Questions??