Different Spirals of Sameness: A Study of Content Sharing in Mainstream and Alternative Media

Benjamin D. Horne, Jeppe Nørregaard, and Sibel Adali

Rensselaer Polytechnic Institute, Technical University of Denmark
Primary focus of this work is on content sharing in alternative media

- Content sharing - one news producer publishes (or copies) an article published by another source
- While it is a fairly common practice in MSM, it is a lesser studied tactic by disinformation producers to get their message across
- Just as bot-driven misinformation, content sharing can be used to:
  - Make particular stories or narratives seem more important,
  - More widely reported,
  - Thus, more credible.
- If we can understand the dynamics of this tactic, we can use the behavior of these producers to better inform readers or algorithms.
Starbird et al. 2018 “Ecosystem or Echo-System?”

- First look at content sharing in alternative media, as a malicious tactic
- The paper focused on spreading alternative narratives about Syrian Civil Defense
- Build an undirected network of shared content to illustrate the narratives spread
- Used a mix-method approach to analyze the data
● “[G]overnment-funded media and geopolitical think tanks as source content for anti-White Helmets narratives.”

● Alternative media ecosystem often shared explicit critiques of MSM

● “[S]mall set of websites and authors generating content that is spread across diverse sites”

Similar exploratory study across more time and events

- We live scrape news sites RSS feeds for **10 months** twice a day, everyday (Feb 2018 to November 2018)
- This live scraping allows us to get almost every published article by each source
- A broad spectrum of sources are collected, including both mainstream and alternative sources (**194 total sources**)
- Focuses broadly on politics, but covers many events
- In total we analyze **713K news articles**

Come to our poster session to here more about the data!
Using this data, we construct a verbatim copy network

1. TFIDF matrix for each 5 day period in the dataset, for each pair of article vectors, compute the cosine similarity, choose article pairs with cosine similarity of 0.85.
2. For each pair, we order them by the UTC timestamp, as to create directed edges from the original article to the copied article.
3. We manually verify pairs, to ensure there are not mistakes due to incorrect timestamps
Content sharing communities represent distinct parts of the media
Echo Chambers & Content Mixing

- O: Russian/conspiracy community
- Y: Right-wing/conspiracy community
- G: U.S. mainstream community
- M: Left-wing blog community
- C: U.K. mainstream community

<table>
<thead>
<tr>
<th></th>
<th>O</th>
<th>Y</th>
<th>G</th>
<th>M</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>13395</td>
<td>1762</td>
<td>484</td>
<td>102</td>
<td>334</td>
</tr>
<tr>
<td>Y</td>
<td>2133</td>
<td>2477</td>
<td>225</td>
<td>872</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>5405</td>
<td>66</td>
<td>1480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>266</td>
<td></td>
<td>323</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>7448</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• 2038 of the 2477 articles copied from the green to yellow are by The Drudge Report
• 160 of these articles are copied by Western Journal.
• Drudge had 138.34M visits in December 2018
• CNN had 497.01M visits in December 2018
Russian/conspiracy community
Right-wing/conspiracy community
U.S. mainstream community
Left-wing blog community
U.K. mainstream community

<table>
<thead>
<tr>
<th></th>
<th>O</th>
<th>Y</th>
<th>G</th>
<th>M</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>13395</td>
<td>1762</td>
<td>484</td>
<td>102</td>
<td>334</td>
</tr>
<tr>
<td>Y</td>
<td>2133</td>
<td>2477</td>
<td>225</td>
<td>872</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>5405</td>
<td>66</td>
<td>1480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>266</td>
<td></td>
<td></td>
<td>323</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7448</td>
</tr>
</tbody>
</table>
Competing Narratives

- News articles are not shared (verbatim or partial) across communities, but the event or topic often is.
- Completely different articles are written based on the same event and are published in several different news communities
- Ultimately creating various narratives around a broad event
- These competing narratives are often repeatedly shared in the alternative news communities, sometimes multiple times by the same source
- Very similar to the competing narrative behavior surrounding the role of the Syrian Civil Defence (Starbird et al. 2018)
Reuters - Kavanaugh accuser agrees to testify in Senate hearing - verbatim copied by OANN. - partially copied by NPR, CBS News, USA Today, Fortune, Mother Jones, Politicus USA, and The Russophile.

Natural News - BOMBSHELL Christine Blasey Fords letter to Sen Dianne Feinstein revealed to be a total FAKE - copied verbatim by The D.C. Clothesline and The Russophile

Natural News - Kavanaugh accuser Christine Blasey Ford ran mass hypnotic inductions of psychiatric subjects - copied verbatim by The Russophile - copied partially by HumansAreFree.


Hot Air - Hmmm Ford hires former Clinton Biden adviser as potential hearing shpapa as attorneys bargain for - copied partially by The Gateway Pundit and Sign of the Times (sottnet)

The Gateway Pundit - Far Left Activist and Kavanaugh Accuser Christine Blasey Ford Spotted at Antitrump March in LA VID - copied partially by The Political Insider and True Pundit


Counter-narratives (Justifying long-standing narratives)

● The Guardian published a story alleging that Paul Manafort, former campaign manager to President Donald Trump, held a secret meeting with Julian Assange, the founder of Wikileaks, inside the Ecuadorian embassy.

● The story was criticized by other well reputed sources for:
  ○ Relying on anonymous sources,
  ○ Not providing any verifiable details
  ○ Being unbelievable given the high level of surveillance in the area surrounding the embassy

● The story was edited by The Guardian multiple times within five hours, weakening the language surrounding the claims

● The story remains unverified

● Yet, The Guardian has not retracted the article or demonstrated any further investigation to verify the report

Natural News - The Guardian caught publishing completely fake news - copied partially by The Duran


The Gateway Pundit - IT WAS A HOAX Guardian Report Blows Up Manafort Passport Shows NO UK TRIPS Never Met with Assange


21stCenturyWire - Wikileaks Rips Guardians Manafort-Assange Report by Serial Fabricator Offers Million Dollar Challenge - copied verbatim by The Russophile

Sputnik - As Guardians Manafort-Assange Story Exposed as Fake Ex-CIA Agent Blames Russia

True Pundit - The Guardian Faceplants As Manaforts Passport Stamps Dont Match Fabricated Assange Story

Daily Stormer - Manafort DID NOT Visit Julian Assange The Guardian Made it All Up

LewRockwell - The Assange-Manafort Fabricated Story - copied verbatim by The Russophile and Sign of The Times (sott.net).

Small breach of journalistic standards can make it uncertain who to trust

- At the surface level, these articles seem to be pushing the standard conspiracy narrative that “the mainstream media is the fake media,”
- Ultimately takes power away from the proper news and gives power to conspiracy theorist
- Due to limited attention and information overload, this small breach may be enough to erode trust
In Conclusion

1. We find that content sharing happens in tightly formed communities, and these communities represent relatively homogeneous portions of the media landscape.

2. We discover mainstream and conspiracy content mixing by several highly read sources in the right-wing community (ex. Drudge Report).

3. We find that alternative news sources repeatedly share content about competing contemporaneous narratives.
Thanks to my Co-authors Jeppe and Sibel

I am on the job market!

Twitter: @benjamindhorne
Extending to Partial Content Sharing

- Extend to partial content sharing by utilizing methods from plagiarism detection
- “Winnowing” - combination of hashing and windowing to create fingerprints for text (Schleimer, Wilkerson, and Aiken 2003)
- This method is only used to extend paths in qualitative analysis, not build a new network
Content sharing among journalistic organizations is well known

● Long line of work in communications, typically focused on imitation in order to meet demand and be competitive

● This imitation behavior has been discussed as early as 1955, “many newspapers feature the same news stories atop their front pages” (Breed 1955)

● Various reasons for this behavior have been argued:
  ○ Popularity of the internet
  ○ Reading news during the work day rather than the ends of days

● Primary concern: news has become homogeneous and significantly less diverse, representing less views and covering less events
Today this homogenized view of the news has been complicated by “alternative” media

- The rise of false, hyper-partisan, and propagandist news producers has created a media environment where:
  - Competing narratives around the same event (Starbird 2017)
  - No gatekeepers to curate quality information (Reese et al. 2009; Allcott and Gentzkow 2017)
  - Add 1 more here
- More diverse news, but lower quality
- Due to this, research has shifted to
  - Detecting low quality information
  - Understanding tactics used to spread low quality information (headline structure, social bots, etc.)
We characterize these communities by using multiple expert labeling systems

1. Credibility -
   a. **NewsGuard** - a group of trained journalists to assess credibility and transparency of news websites. NewsGuard assesses nine journalistic criteria which are combined into a good or bad
   b. **Media Bias/Fact Check** - analyzes news sources to determine their credibility using trained team for factuality
   c. We combine their factual-reporting score with NewsGuard’s credibility label, for a final label of source reliability

2. Political Leaning -
   a. **Media Bias/Fact Check** - MBFC provides a descriptive label for sites, which often includes the source’s political bias across the political spectrum
   b. **Buzzfeed hyperpartisan list** - list hyper-partisan sites with binary left or right label
   c. We aggregate a bias score for each source by normalizing each rating from -1 (left) to 1 (right)

3. Country - manually check the country of origin of each source if known
Community Characteristics:
89% U.S. sources
40% Left, 34% Center
74% Credible
Community Characteristics:
63% U.K. sources
31% Center, 37% Unknown
14% Credible, 86% Unknown
Community Characteristics:
94% U.S. sources
31% Center, 37% Unknown
47% Credible, 53% Unknown

Small size is not due to imbalance in data, but less verbatim copying done by this community
Community Characteristics:
94% U.S. sources
70% Right, 15% Unknown
27% Credible, 57% Unknown
Community Characteristics:
54% U.S. sources, 21% Russian
10% Right, 62% Unknown
8% Credible, 87% Unknown
Using these communities, we perform a mixed-method analysis

1. Echo Chambers
2. Content Mixing
3. Competing Narratives
4. Counter Narratives