Fake news on Twitter

Lisa Friedland, Kenny Joseph, Nir Grinberg, David Lazer
Northeastern University
Case study of a fake news pipeline

**Step 1:** Wikileaks acquires hacked emails from John Podesta

**Step 2:** Wikileaks publishes emails

---

Dear Tony,

I am so looking forward to the Spirit Cooking dinner at my place. Let me know if your brother is joining?
Case study (cont.)

**Step 3a:** Fake news derived from leaks

**Step 3b:** Diffusion & misinformation campaigns
Case study (cont.)

#SpiritCooking texted a couple Hillary voters to read this hashtag and they are changing vote to Trump. Only one that can stop this. Sick

Step 4: Potential Impacts on voters
Flow of fake news about Podesta

Websites
- thegatewaypundit.com
- truthfeed.com
- infowars.com
- zerohedge.com
- theamericanconservative.com
- trunck.com
dcjobwise.com
endingtheed.com
freedomcom.com
worriedvague.org
yournewsire.com

Introduced to Twitter
- PrisonPlanet
- WikiLeaks
- DarronVirgil
- JillBare
- TrumpRomn
- Trump
- AnnBarak
- YvetsJames
- MikeCoulter
- MOVETOFRW
- Nickolas
- PastorScott
- HenryJackson
- rejack

Panel Members
- @AA2017
- @realDonaldTrump
- @realDonaldTrump
- @realDonaldTrump
- @realDonaldTrump
- @realDonaldTrump
- @realDonaldTrump
- @realDonaldTrump
- @realDonaldTrump
- @realDonaldTrump
- @realDonaldTrump
Fake news on Twitter

1. Data: Americans sharing URLs on Twitter
2. How prevalent was fake news on Twitter?
3. Who shared fake news?
4. How did it spread through Twitter?
Studying **humans** on Twitter

If we want to know how fake news affected the election, look at voters!
### Voter registrations

<table>
<thead>
<tr>
<th>Name</th>
<th>State</th>
<th>Party</th>
<th>Age</th>
<th>Match to Twitter?</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kenny Joseph</td>
<td>MA</td>
<td>D</td>
<td>52</td>
<td>Yes!</td>
<td></td>
</tr>
<tr>
<td>Nir Grinberg</td>
<td>NY</td>
<td>D</td>
<td>19</td>
<td>No, not unique</td>
<td></td>
</tr>
<tr>
<td>Nir Grinberg</td>
<td>NY</td>
<td>R</td>
<td>25</td>
<td>No, not unique</td>
<td></td>
</tr>
<tr>
<td>Amelia Tzray</td>
<td>CA</td>
<td>…</td>
<td>…</td>
<td>No, not on Twitter</td>
<td></td>
</tr>
</tbody>
</table>

#### Twitter universe

- Matching 2 million registered voters:
  - No matches on Twitter
  - >1 match
  - Exactly 1 match...
    - 10% blank location field
    - 2% location not parsable
    - 1% correct U.S. state: 22,861
  - No matches to correct state

Result: panel of ~22k U.S. voters on Twitter
We focus on ~100K “political URL” tweets that:

a) mention Clinton or Trump
b) were sent in 2016
c) have URLs (not back to Twitter)
Defining fake news

• Compiled list of 449 “fake news” sources (websites)
  • Aggregation of existing lists from Hoaxy, BuzzFeed, Media Bias Fact Check, ...

• To validate coverage, manually examine popular URLs from fake and uncategorized sources
  • Does it have the veneer of legitimate news?
  • Is the content truthful?
Exploring fake and real news

• Much political content (both real and fake) on the web resembles clickbait or blogs
  • Casual writing style, hyperlinks to story sources, highly partisan

• We see a spectrum of truthfulness. Often hard to distinguish among:
  • Clickbait → real
  • Misleading → fake
  • Conspiracy → fake
  • Fictional → fake
  (Easiest to recognize, but rare)
How well do the lists capture fake news?

- On fake news lists: fairly “sketchy” content
- Not on lists: still 4% of URLs are sketchy
- The categorizations can be improved, but these are reasonable for a first pass

Includes "Misleading," "Conspiracy," and "Fictional" categories
Fake news on Twitter

1. Data: Americans sharing URLs on Twitter
2. How prevalent was fake news on Twitter?
3. Who shared fake news?
4. How did it spread through Twitter?
Fake news was more prevalent than you might expect

- 5,503 out of 95,660 (5.8%) links in 2016 were to fake news
- Most prominent in months leading up to election
It was even prominent compared to individual “real” news sources
# Ranking of popularity of fake news

<table>
<thead>
<tr>
<th>Rank</th>
<th>“Website”</th>
<th>% Republicans w/ ≥ 1 URL from site</th>
<th>% Democrats w/ ≥ 1 URL from site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NY Times</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>Youtube</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>Washington Post</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>CNN</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>Facebook</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Politico</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>The Hill</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>Fake News</td>
<td><strong>11</strong></td>
<td><strong>7</strong></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Breitbart</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>
The most popular fake news sites were right leaning.

Top 50 overall (▲) & fake news (⬤) sites by % Dems, Reps who tweeted out at least 1 URL from site.
Fake news on Twitter

1. Data: Americans sharing URLs on Twitter
2. How prevalent was fake news on Twitter?
3. Who shared fake news?
4. How did it spread through Twitter?
Fake news sharing was heavy-tailed

- 4,790 users (of 22K) shared at least 1 "political URL"
- Of those, 7% (348) shared a fake news URL
- 70% of fake news was shared by a set of 15 people!
  - More evenly distributed than Fox, similar to Breitbart
  - Less evenly distributed than traditional news outlets, social media
Who are the super sharers?

- Real people
- Highly active users
- Use automation at times
- Super-sharers of fake news tend to be super-sharers of news in general
Fake news on Twitter

1. Data: Americans sharing URLs on Twitter
2. How prevalent was fake news on Twitter?
3. Who shared fake news?
4. How did fake news spread through Twitter?
Fake news tweets and retweets

- Fake news dominated by 3 websites, 15 super-sharers
- Many introducers
- Panel members seem to retweet without regard to the introducer
Retweets: Fake news vs. all political URLs
Conclusions

• Fake news was among the most **popular** news content shared on Twitter in October 2016

• **A small number of websites** were main producers of fake news content (subject to our definitions)

• Most fake news (70%) was shared by **a small number of super-sharers**
  • Only 15 in a 22K sample (of non-bots)
  • Had “cyborg tendencies”
  • Other 30% shared by right leaning, older people

• Channels where fake news was shared:
  • Often cross-posted from Facebook
  • Different, highly partisan, intermediate sources
  • Voters who shared fake news did so without much regard to specific primary or intermediate sources