Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter?

COMPROP DATA MEMO 2017.1 / 26 March 2017

Philip N. Howard
Oxford University
philip.howard@oii.ox.ac.uk
@pnhoward

Gillian Bolsover
Oxford University
gillian.bolsover@oii.ox.ac.uk

Bence Kollanyi
Oxford University
bence.kollanyi@oii.ox.ac.uk
@bencekollanyi

Samantha Bradshaw
Oxford University
samantha.bradshaw@oii.ox.ac.uk
@sbradshaww

Lisa-Maria Neudert
Oxford University
lisa-maria.neudert@oii.ox.ac.uk
@lmneudert

ABSTRACT

Computational propaganda distributes large amounts of misinformation about politics and public policy over social media platforms. The combination of automation and propaganda can significantly impact public opinion during important policy debates, elections, and political crises. We collected data on automation and junk news using major hashtags related to politics in the state of Michigan in the lead up to the 2016 US Presidential Election. (1) In Michigan, conversation about politics over Twitter mirrored the national trends in that Trump-related hashtags were used more than twice as often as Clinton-related hashtags. (2) Social media users in Michigan shared a lot of political content, but the amount of professionally researched political news and information was consistently smaller than the amount of extremist, sensationalist, conspiratorial, masked commentary, fake news and other forms of junk news. (3) Not only did such junk news "outperform" real news, but the proportion of professional news content being shared hit its lowest point the day before the election.

COMPUTATIONAL PROPAGANDA AND THE 2016 ELECTION

Social media plays an important role in the circulation of ideas about public policy and politics. Political actors and governments worldwide are employing both people and algorithms to shape public life.[1,2] Bots are software intended to perform simple, repetitive, and robotic tasks. They can perform legitimate tasks on social media like delivering news and information—real news as well as junk—or undertake malicious activities like spamming, harassment and hate speech. Whatever their uses, bots on social media platforms are able to rapidly deploy messages, replicate themselves, and pass as human users. They are also a pernicious means of spreading junk news over social networks of family and friends.

Computational propaganda flourished during the 2016 US Presidential Election. There were numerous examples of misinformation distributed online with the intention of misleading voters or simply earning a profit. Multiple media reports have investigated how "fake news" may have propelled Donald J. Trump to victory.[3–5] In Michigan, pre-election polls showed the two presidential candidates relatively close in voter support, making this state an ideal case in which to study the prevalence and distribution of fake news during the Presidential Election. What kinds of political news and information were circulating over social media among voters in Michigan? How much of it was extremist, sensationalist, conspiratorial, masked commentary, fake, or some other form of junk news?

SOCIAL MEDIA AND JUNK NEWS

Junk news, widely distributed over social media platforms, can in many cases be considered to be a form of computational propaganda. Social media platforms have served significant volumes of fake, sensational, and other forms of junk news at sensitive moments in public life, though most platforms reveal little about how much of this content there is or what its impact on users may be. The World Economic Forum recently identified the rapid spread of misinformation online as among the top 10 perils to society.[6] Prior research has found that social media favors sensationalist content, regardless of whether the content has been fact checked or is from a reliable source.[7] When junk news is backed by automation, either through dissemination algorithms that the platform operators cannot fully explain or through political bots that promote content in a preprogrammed way, political actors have a powerful set of tools for computational propaganda.[8] Both state and non-state political actors deliberately manipulate and amplify non-factual information online.

Fake news websites deliberately publish misleading, deceptive or incorrect information purporting to be real news for political, economic or cultural.[9] These sites often rely on social media to attract web traffic and drive engagement. Both fake news websites and political bots are crucial tools in digital propaganda attacks—they aim to influence conversations, demobilize opposition and generate false support.

Since the UK's Brexit Referendum and the US Presidential Election of 2016, fake news has been under much scrutiny for degrading public knowledge of important trends and issues. However, very little is yet known about the prevalence of fake news during political events and rarely are single examples of fake news set in context within a larger media ecosystem of sources.

SAMPLING AND METHOD

Our analysis is based on a dataset of approximately 22m tweets collected between November 1-11, 2016, that contained hashtags related to politics and the election in the US. Our previous analyses have been based on samples of political conversation, over Twitter, that used hashtags relevant to the US election as a whole. However, in the subsample of political conversation over Twitter reported here, we select users who provided a city and state name from Michigan. Michigan was chosen because it was a key battleground state where public support was evenly split between both candidates right up to Election Day.

Within our initial sample of 22m tweets, 138,686 were from users who provided location information from the state of Michigan through the manual input of a city or state in the location field of their profiles. These tweets and associated data were collected from Twitter's public API. The platform's precise sampling method is not known, but the company itself reports that the data available through the Streaming API is at most one percent of the overall global public communication on Twitter at any given time.[10] Tweets were selected based on a list of keywords and hashtags associated with the US election and tweets were collected from the API that (1) contained at least one of the relevant hashtags; (2) contained the keyword or hashtag in the text of a link, such as a news article, shared in that tweet; (3) were a retweet of a message that contained the keyword or hashtag in the original message; or (4) quoted tweets in which the keyword or hashtag was included but in which the original text was not included and Twitter used a URL to refer to the original tweet.

To assess the prevalence of hashtags and keywords in the dataset, we count these in a straightforward way. Each tweet was automatically coded and counted if it contained one of the specific hashtags that were being followed. If the same hashtag was used multiple times in a tweet, this method counted that tweet only once. If a tweet contained more than one selected hashtag, it was credited to all the relevant hashtag categories.

To assess the prevalence of different types of information sources being shared on social media, we determined the source of each of the URLs in the dataset; 25,339 of the 138,686 tweets by Michigan-based users about the US election between 1 and 11 November contained a URL. Some 7.8% of the sources were no longer available. For the rest of the URLs, each source that was shared more than three times was catalogued, and many sources that were easy to identify but shared less often than that were also catalogued. Effectively this typology is built on successful cataloguing of 99% of the content shared, with the remainder being single URLs shared only a few times or otherwise broken in some way.

Our two-stage coding process involved developing an initial, grounded coding scheme and running it as a kind of pilot study of a first wave of some 3,500 URLs. We then revised our categories and definitions and recoded the complete dataset according to the categories defined below.

The limitations of this methodology should be noted. Tweets about the US Presidential Election by individuals in Michigan who did not use one of these hashtags would not have been captured. Tweets from people who used these hashtags but were tweeting about something else, would be captured in this sample. The coding of source types was derived from the dataset and is not intended to be a comprehensive list of all types of information providers. The overall percentages of different information sources are intended as an indication of the information environment surrounding the Presidential Election in Michigan to stimulate further research and conversation.

FINDINGS AND ANALYSIS

This sample allows us to draw some conclusions about the character and process of political conversation over Twitter during the election, particularly as it relates to Michigan voters and the circulation of different kinds of news and political information among voters in Michigan.

Comparing the Candidates on Twitter in Michigan. Table 1 reveals that 138,686 tweets were posted in Michigan about politics and the election. This table shows that the overall volume of tweets using pro-Trump hashtags (56.7 percent), was much greater than the volume of tweets containing only hashtags associated with the Clinton campaign (20.3 percent). The overall volume of tweets using neutral election-related hashtags (13.4 percent) was also significantly lower compared with those using pro-Trump hashtags.

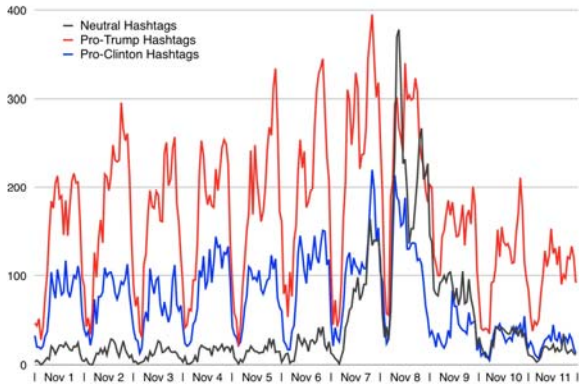
Figure 1 shows the rhythm of this traffic over the sample period. It reveals that Trump-related traffic significantly outpaced Clinton-related traffic over the course of the eleven-day period. Neutrally tagged content—usually about having successfully voted—peaked on Election Day and surpassed the volume of traffic simply about one candidate or the other.

In previous memos we attempted to catalogue the users with high levels of automation behind their accounts, but for this state-specific sub-sample no accounts generated more than our usual threshold for identifying automation. A close look revealed that only 2% of the platforms used to send Twitter traffic were known sources of bots—the rest were platforms that either supported human users or offered very human-like levels of automation. It is

Table 1: Twitter Conversation about Michigan Politics around Voting Day, 2016		
	N	%
Pro-Trump	78,662	56.7
Pro-Clinton	28,074	20.3
Neutral	18,613	13.4
Trump-Neutral	2,949	2.1
Clinton-Neutral	1,464	1.1
Trump-Clinton	8,361	6.0
Trump-Clinton-Neutral	563	0.4
Total	138,686	100.0

Source: Authors' calculations from data sampled 1-11/11/16.
Note: Pro-Trump hashtags include #AmericaFirst, #benghazi, #CrookedHillary, #DrainTheSwamp, #lockherup, #maga3x, #MAGA, #MakeAmericaGreatAgain, #NeverHillary, #PodestaEmails, #projectveritas, #riggedelection, #tcot, #Trump2016, #Trump, #TrumpPence16, #TrumpTrain, #VoterFraud, #votetrump, #wakeupamerica; pro-Clinton hashtags include #Clinton, #ClintonKaine16, #democrats, #dems, #dnc, #dumptrump, #factcheck, #hillary2016, #Hillary, #HillaryClinton, #hillarysupporter, #hrc, #ImWithHer, #LastTimeTrumpPaidTaxes, #NeverTrump, #OHHillYes, #p2, #strongertogether, #trumptape, #uniteblue; neutral hashtags include #Election2016, #Elections2016, #uselections, #uselection, #earlyvote, #iVoted, #Potus.

Figure 1: Hourly Twitter Traffic in Michigan, by Candidate



Source: Authors' calculations from data sampled 1-11/11/16.
Note: This figure is based on the hashtags used in the tweets.

possible that some of these platforms are obscuring accounts that really are highly automated. But we believe humans are more likely to volunteer geolocation information than bots, and that high levels of automation are easier to detect during broad, nation-wide political conversations than in more local subsamples of traffic. In other words, bots may be less active or harder to detect in state-level, subnational political conversations, and more active or easier to detect at the national level.

What Were Michiganders Sharing? Table 2 catalogues the different kinds of URLs being shared in election related Tweets by users located in Michigan. The sources of these URLs were coded as one of sixteen different types according to the following coding scheme:

- Professional News Outlets.
  - o Major News Brands. This is political news and information by major outlets that display the qualities of professional journalism, with fact-checking and credible standards of production. They provide clear information about real authors, editors, publishers and owners, and the content is clearly produced by an organization with a reputation for professional journalism. This content comes from significant, branded news organizations, including any locally affiliated broadcasters.
  - o Minor News Brands. As above, but this content comes from small news organizations or startups that display evidence of organization, resources, and professionalized output that distinguishes between fact-checked news and commentary.

- Professional Political Content
  - o Government. These links are to the websites of branches of government or public agencies.
  - o Experts. This content takes the form of white papers, policy papers, or scholarship from researchers based at universities, think tanks or other research organizations.
  - o Political Party or Candidate. These links are to official content produced by a political party or candidate campaign.

- Other Political News and Information
  - o Junk News. This content includes various forms of propaganda and ideologically extreme, hyper-partisan, or conspiratorial political news and information. Much of this content is deliberately produced false reporting. It seeks to persuade readers about the moral virtues or failings of organizations, causes or people and presents commentary as a news product. This content is produced by organizations that do not employ professional journalists, and the content uses attention grabbing techniques, lots of pictures, moving images, excessive capitalization, ad hominem attacks, emotionally charged words and pictures, unsafe generalizations and other logical fallacies.
  - o WikiLeaks. Tweets with these links usually offer unverified claims and the suggestion that WikiLeaks.org provides evidence.
  - o Citizen, Civic, or Civil Society. Links to content produced by independent citizens, civic groups, or civil society organizations. Blogs and websites dedicated to citizen journalism, citizen-generated petitions, personal activism, and other forms of civic expression that display originality and creation more than curation or aggregation.
  - o Humor and Entertainment. Content that involves political jokes, sketch comedy, political art or lifestyle- or entertainment-focused coverage.
  - o Religion. Links to political news and information with distinctly religious themes and faith-based editorializing presented as political news or information.
  - o Russia. This content was produced by known Russian sources of political news and information.
  - o Other Political Content. Myriad other kinds of political content, including portals like AOL and Yahoo! that do not themselves have editorial policies or news content, survey providers, and political documentary movies.

- Other
  - o Social Media Platforms. Links that simply refer to other social media platforms, such as Facebook or Instagram. If the content at the ultimate destination could be attributed to another source, it is.
  - o Other Non-Political. Sites that do not appear to be providing information but that were, nevertheless, shared in tweets using election-related hashtags.

- No Longer Available. These links were shared during the sample period, but the content being linked to has since been removed. If some evidence from an author or title field, or the text used in a UR could be attributed to another source, it is.

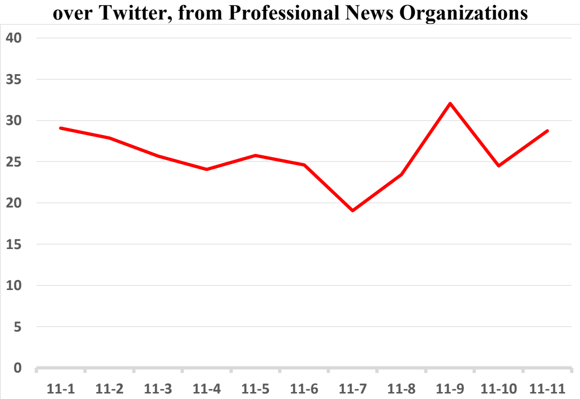
Table 2 presents the findings of this grounded catalogue of content. Overall, 25.9% of the political news and information being shared by Twitter users in Michigan came from professional news organizations. Links to content produced by

Table 2: What Political News and Information Was Michigan Sharing Over Twitter?

Type of Source	*N*	*%*	*N*	*%*
Professional News Content				
Major News Brands	4,684	73.1		
Minor News Brands	1,724	26.9		
Subtotal	6,408	100.0	6,408	25.9
Professional Political Content				
Political Party or Candidate	658	78.1		
Experts	121	14.4		
Government	64	7.6		
Subtotal	843	100.0	843	3.4
Other Political News and Information				
Junk News	6,469	54.5		
Humor or Entertainment	2,024	17.1		
WikiLeaks	1,186	10.0		
Citizen or Civil Society	908	7.7		
Other Political	700	5.9		
Political Merchandise	249	2.1		
Russia	211	1.8		
Religion	121	1.0		
Subtotal	11,868	100.0	11,868	47.9
Other				
Social Media Platform	3,038	81.5		
Other Non-Political	689	18.5		
Subtotal	3,727	100.0	3,727	15.0
No Longer Available	1,937		1,937	7.8
Total	24,783		24,783	100.0

Source: Authors' calculations from data sampled 1-11/11/16.

Figure 2: Percent of Political News and Information, Shared over Twitter, from Professional News Organizations



Source: Authors' calculations from data sampled 1-11/11/16.

government agencies, political parties and candidates, or experts, altogether added up to just 3.4% of the total. Indeed, only small fractions of the content being shared originated with the political parties, candidates, civil society groups, universities or public agencies.

The category of "Other Political News and Information" includes many different kinds of content. The number of links to junk news alone is roughly equivalent to the number of links to professionally researched journalism. But other forms of questionable sources abound with conspiracy videos, links to unverified WikiLeaks content, and other political content that is too complex to fit into a parsimonious typology. Two things should be noted across categories. First, the proportion of professional to junk news is roughly one-to-one. Second, when the amount of junk news is added to the number of links to unverified WikiLeaks content, and Russian-origin news stories, it appears that fully 46.5 percent of all the content that is presented as news and information about politics and the election is of an untrustworthy provenance or falls under the definition of propaganda based on its use of language and emotional appeals.

Figure 2 provides the percent of political news and information that was produced by professional news organizations and shared, day-by-day, in the lead up to the election. While the overall volume of content being shared increased each day closer to the election, the overall proportion of that content coming from professional news organizations actually diminished. On the day before election day, November 7th 2016, the volume of all the other forms of political news and information peaked and the relative proportion of professional news and information was at its lowest.

CONCLUSIONS

The internet has long been used both for political activism and social control.[11] Political conversation from and about Michigan mirrored that of the nation in that Trump's presence on Twitter was consistently more than twice that of Clinton's, and the use of neutral hashtags for tweeting about politics was minor. User sentiment from this sample is different from that of the voters who responded to public opinion polls, however, which showed support for Trump and Clinton to be relatively equal right up to election day.

The term "fake news" is difficult to operationalize, so our grounded typology reflects the diversity of organizations behind the content that was circulated over Twitter by people in Michigan. Social media users in Michigan shared many links to political news and information, but junk news, characterized by ideological extremism, misinformation and the intention to persuade readers to respect or hate a candidate or policy based on emotional appeals, was just as, if not more, prevalent than the amount of information produced by professional news organizations. Not only did this computational propaganda "outperform" real news in Michigan in the lead up to the presidential election, but the proportion of professional news content being shared hit its lowest point the day before the election.

ABOUT THE PROJECT

The Project on Computational Propaganda (www.politicalbots.org) involves international, and interdisciplinary, researchers in the investigation of the impact of automated scripts—computational propaganda—on public life. *Data Memos* are designed to present quick snapshots of analysis on current events in a short format. They reflect methodological experience and considered analysis, but have not been peer-reviewed. *Working Papers* present deeper analysis and extended arguments that have been collegially reviewed and that engage with public issues. The Project's articles, book chapters and books are significant manuscripts that have been through peer review and formally published.

ACKNOWLEDGMENTS AND DISCLOSURES

The authors gratefully acknowledge the support of the European Research Council, "Computational Propaganda: Investigating the Impact of Algorithms and Bots on Political Discourse in Europe," Proposal 648311, 2015-2020, Philip N. Howard, Principal Investigator. Project activities were approved by the University of Oxford's Research Ethics Committee. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the European Research Council.

REFERENCES

1. Howard, P. N. & Kollanyi, B. Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum. *ArXiv160606356 Phys.* (2016).
2. Forelle, M., Howard, P., Monroy-Hernández, A. & Savage, S. Political Bots and the Manipulation of Public Opinion in Venezuela. *ArXiv150707109 Phys.* (2015).
3. Parkinson, H. J. Click and elect: how fake news helped Donald Trump win a real election. *The Guardian* (2016).
4. Read, M. Donald Trump Won Because of Facebook. *New York Magazine* (2016).
5. Dewey, C. Facebook Fake-News writer: 'I Think Donald Trump is in the White House Because of Me'. *The Washington Post* (2016).
6. World Economic Forum. 10. The Rapid Spread of Misinformation Online. *Outlook on the Global Agenda 2014* (2014).
7. Vicario, M. D. *et al.* The Spreading of Misinformation Online. *Proc. Natl. Acad. Sci.* **113,** 554–559 (2016).
8. Kümpel, A. S., Karnowski, V. & Keyling, T. News Sharing in Social Media: A Review of Current Research on News Sharing Users, Content, and Networks. *Soc. Media Soc.* **1,** 2056305115610141 (2015).
9. Howard, P. N. Digitizing the social contract: Producing American political culture in the age of new media. *Commun. Rev.* **6,** 213–245 (2003).
10. Morstatter, F., Pfeffer, J., Liu, H. & Carley, K. M. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. *ArXiv13065204 Phys.* (2013).
11. Howard, P. N. *Pax Technica: How the Internet of Things May Set Us Free or Lock Us Up*. (Yale, 2015).