Junk News and Bots during the French Presidential Election: What Are French Voters Sharing Over Twitter?

COMPROP DATA MEMO 2017.3 / 22 APRIL 2017

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

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

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

Clementine Desigaud
Oxford University
clementine.desigaud@oii.ox.ac.uk
@cmdesigaud

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

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 Twitter data on bot activity and junk news using a set of hashtags related to the French Presidential Elections for a week in March 2017. (1) Content about Macron tended to dominate the traffic on Twitter, but highly automated accounts occasionally generated large amounts of traffic about Hamon. (2) These automated accounts generated a small amount of content about French politics, though this amount is increasing over time. (3) Social media users in France shared many links to high quality political news and information, roughly at ratio of 2 links to professionally produced news for every link to other kinds of sources. (4) In comparison to our study of similar trends in the US and Germany, we find that French users are sharing better quality information than what US users and German users share.

SOCIAL MEDIA AND AUTOMATION

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, one of the key battleground states, junk news was shared just as widely as professional news in the days leading up to the election.[1] There is growing evidence that social-media platforms support campaigns of political misinformation campaigns on a global scale. News and media outlets have covered stories about the alt-right and other extremist movements gaining momentum in France.

JUNK NEWS AND AUTOMATION

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. What kinds of political news and information are circulating over social media among voters in France? How much of it is high-quality, professional news, and how much content is extremist, sensationalist, conspiratorial, masked commentary, fake, or some other form of junk news? Is there evidence of a similar widespread disinformation campaign occurring in France?

SAMPLING AND METHOD

Our dataset contains approximately 842K tweets collected between 13-19 March 2017, using a combination of hashtags associated with the primary Presidential candidates.

Twitter provides free access to a sample of the public tweets posted on the platform. 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 any given time.[10] In order to get the most complete and relevant data set, we consulted with country experts and used our pilot study data to identify relevant hashtags. The following hashtags were selected for our analysis: #AuNomDuPeuple, #benoithamon, #electionpresidentielle, #elections2017, #Fillion, #fillion2017, #FillonGate, #FrancoisFillon, #FrançoisFillon, #frontnational, #Hamon, #Hamon2017, #LePen, #lesrepublicains, #Macron, #Macron2017, #Marine, #Marine2017, #MarineLePen, #melenchon, #Melenchon2017, #MLP, #MLP2017, #Mélenchon, #Mélenchon2017, #PenelopeGate, #Presidentielle, #Presidentielle2017, #Presidentielles, #Présidentielle, #présidentielle2017, #Présidentielles. Parliamentary and multi-party systems tend to have a greater variety of hashtags related to particular candidates and important political issues. Thus, our sampling strategy might have missed some additional minor hashtags that refer to small or short lived conversations about particular people or issues, including tweets that may not have used our identified hashtags at all. The programming of the data collection and most of the analysis were done by using the R software environment developed for statistical computing and graphics.

Selecting tweets based on hashtags has the advantage of capturing the content most likely to be about this important political event. The streaming API yields (1) tweets which contain the keyword or the hashtag; (2) tweets with a link to a web source, such as a news article, where the URL or the title of the web source includes the keyword or hashtag; (3) retweets that contain a message's original text, wherein the keyword or hashtag is used either in the retweet or in the original tweet; and (4) quote tweets where the original text is not included but Twitter uses a URL to refer to the original tweet.

Our method counted tweets with selected hashtags in a simple manner. Each tweet was 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 still counted that tweet only once. If a tweet contained more than one selected hashtag, it was credited to all the relevant hashtag categories.

Contributions using none of these hashtags were not captured in this data set. It is also possible that users who used one or more of these hashtags, but were not discussing the election, had their tweet captured. Moreover, if people tweeted about the election, but did not use one of these hashtags or identify a candidate account, their contributions were not analyzed here.

After determining how often each candidate was being discussed on Twitter, the next step was to determine what information was being shared as political news and information. From our dataset of 842,146 tweets, we selected all of the tweets that contained URLs. Between 13-19 March, Twitter users in France shared 88,755 links on the platform. URLs that pointed towards another tweet were removed from our sample, as most of these tweets were generated automatically by Twitter when someone quotes a tweet. If Twitter users shared more than one URL in their tweet, only the first URL was saved. We then generated a random 10 percent sample of dataset. 8,876 URLs were selected randomly using a python script. We then removed duplicate URLs from our sample to classify each URL according to our classification system (outline below). The classification of each URL was carried out by a team of coders fluent in the French language and familiar with the media landscape. They worked together over a period of two days to ensure consistency while coding URLs. Once each unique URL was coded, we expanded the coding to the duplicate URLs to complete the coding for our random 10 percent sample.

FINDINGS AND ANALYSIS

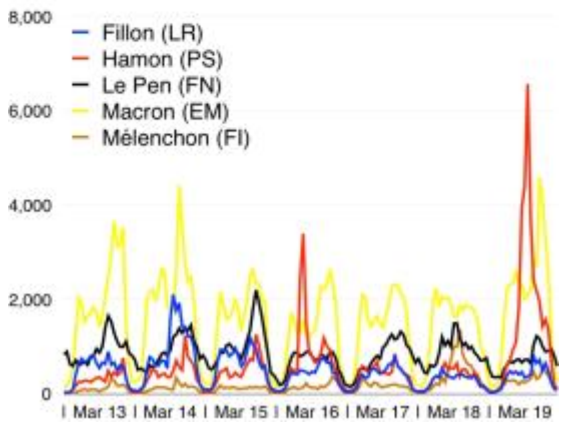
Twitter conversation about French politics can be analyzed in terms of the relative use of candidate

Table 1: Twitter Conversation about the French Presidential Candidates

	N	%
François Fillon (LR)	80,760	13.0
Benoît Hamon (PS)	99,898	16.1
Marine Le Pen (FN)	141,352	22.8
Emmanuel Macron (EM)	267,842	43.2
Jean-Luc Mélenchon (FI)	30,521	4.9
Total	620,373	100.0

Source: Authors' calculations from data sampled 03/13-03/19.
Note: Hamon hashtags include #Hamon, #Hamon2017, #BenoitHamon; Fillon hashtags include #Fillon, #Fillon2017, #FrancoisFillon, #FrançoisFillon; Le Pen hashtags include #LePen, #Marine2017, #MarineLePen, #MLP2017; Melenchon hashtags include #Melenchon, #Melenchon2017, #Mélenchon, #Mélenchon2017; Macron hashtags include #Macron, #Macron2017.

Figure 1: Hourly Twitter Conversation about the French Presidential Candidates



Source: Authors' calculations from data sampled 03/13-03/19.
Note: This figure is based on the candidate-specific hashtags used in the tweets.

hashtags, the level of automation, and the kinds of sources for political news and information.

Table 1 and Figure 1 compare the use of candidate hashtags for the sample week in March 2016. Hashtags about Emmanuel Macron appeared most often—43.2 percent of the candidate-specific tweets during the week as a whole. For the most part, traffic about Marine Le Pen is about half that of Macrons. Yet Figure 1 reveals that day to day there were moments when traffic about Benoît Hamon rose to prominence, even though only 16.1 percent of the content overall related to him.

Table 2 and Figure 2 reveal the rhythm of Twitter traffic about the French Election. We define a high level of automation as accounts that post at least 50 times a day on one of the selected hashtags during the data collection period. This detection methodology fails to capture highly automated accounts that are tweeting with lower frequencies. By candidate, it appears that most of the candidates have roughly the same number of highly automated accounts generating traffic about them. By comparatively, these highly automated users generate different proportions of the candidate's traffic. At the high end, 11.4 percent of the Twitter traffic about Hamon is driven by highly automated accounts. At the low end, 4.6 percent of the Twitter traffic about Mélenchon is generated by highly automated accounts. We cannot know who manages these accounts, and we do not analyze the content or emotional valence of particular tweets. So this information alone is insufficient to determine whether the highly automated accounts are run by the campaign to promote a candidate, or run by outsiders to critique the candidate. Figure 2 reveals that the level of automation being used in French political conversations is fairly consistent and that it flows in tandem with human generated content during the natural waking hours of human users. On average, 7.2 percent of the traffic about French politics is generated by the bots we are able to track. An additional sample of French Election information was taken in February, though it is not analyzed in depth here. By comparison the level of automation rose slightly over time, from an average of 6.8 percent. In other words, over the course of the month the number of tweets being generated by highly automated accounts increased slightly.

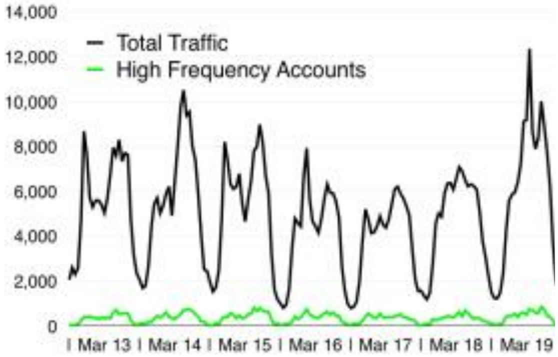
To understand what kinds of political news and information French voters are sharing, we then analyzed the links included in the tweets that also contained relevant hashtags about the French election. Table 3 explains the distribution of content shared by French Twitter users, according to this grounded typology of news platforms and content types.

Table 2: High Frequency Tweeting about the French Election

	N of Tweets	% of Tweets	N of Accounts
François Fillon (LR)	5,955	7.4	105
Benoît Hamon (PS)	11,395	11.4	104
Marine Le Pen (FN)	7,991	5.7	103
Emmanuel Macron (EM)	23,846	8.9	106
Jean-Luc Mélenchon (FI)	1,410	4.6	93

Source: Authors' calculations from data sampled 13-19/03/17. Note: This table is based on the candidate-specific hashtags used in the tweets.

Figure 2: High Frequency Tweeting about the French Election, Hourly



Source: Authors' calculations from data sampled 03/13-03/19. Note: This figure is based on all the hashtags used in the tweets.

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

- Other Political News and Information
  - 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.
  - WikiLeaks. Tweets with these links usually offer unverified claims and the suggestion that WikiLeaks.org provides evidence.
  - 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. Spam is also included in this category.

- Inaccessible
  - o 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.
  - o Language: Links that led to content in foreign language that was neither English nor French, when their affiliation could not be verified through reliable source.

Table 3 reveals that the largest proportion of content being shared by Twitter users interested in French politics comes from professional news organizations. Information from political parties, government agencies and other experts is also being used. Still, 19.6 percent of the content being shared involves other kinds of political news and information. Yet the largest proportion of that content is not junk news but citizen-generated content. The proportion of content posted to personal and organizational blogs is quite large—especially in comparison to Germany or the United States. Only 21.3 percent of this category of political news not produced by traditional journalists or experts actually qualifies as junk.

Table 3: French Political News and Information On Twitter

Type of Source	*N*	*%*	*N*	*%*
Professional News Content				
Major News Brands	3,812	93.0		
Minor News Brands	286	7.0		
Subtotal	4,098	100.0	4,098	46.7
Professional Political Content				
Political Party or Candidate	1,189	86.2		
Government	142	10.3		
Experts	48	3.5		
Subtotal	1,379	100.0	1,379	15.7
Other Political News and Information				
Citizen or Civil Society	725	42.2		
Junk News	366	21.3		
Other Political	337	19.6		
Russia	214	12.5		
Humor or Entertainment	55	3.2		
Religion	15	0.9		
WikiLeaks	6	0.3		
Political Merchandise	-	0.0		
Subtotal	1,718	100	1,718	19.6
Other				
Social Media Platform	879	64.7		
Other Non-Political	480	35.3		
Subtotal	1,359	100	1,359	15.5
Inaccessible				
Language	115	50.2		
No Longer Available	114	49.8		
Subtotal	229	100	229	2.6
Total			8,783	100

Source: Authors' calculations from data sampled 03/13-03/19..

CONCLUSIONS

The internet has long been used both for political activism and social control.[11] 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 tweeting about French politics.

Over time we have been able to compare the consumption of professional news across several countries. French social media users appear to share less junk news than social media users who are actively discussing German politics and the upcoming German election, and political conversations about French politics over social media are not as poisoned as the same kinds of conversations about the US election in 2016. In the days leading up to the US election, we did a close study of junk news consumption among Michigan voters and found a 1:1 ratio between professional news content and junk. In the days leading up to the German Presidential election—admittedly a non-controversial one—we found Germans sharing 4 professionally produced news stories for every 1 link to other kinds of political news and information.

Substantive differences between the qualities of political conversations are evident in other ways. In the US sample, 25.9 percent of all the links being shared led to professional news content and 3.4 of the links led to content from traditional political parties, government agencies, or other experts. In the German sample, 44.9 percent of all the links led to professional news and 13.7 led to expert content. In France, 46.7 percent of all links led to professional news and 15.7 to political parties, government agencies and experts. The people discussing French and German politics over social media tend to use more high quality information sources than those discussing US politics.

Content about Macron tended to dominate the traffic on Twitter, but highly automated accounts occasionally generated large amounts of traffic about

Hamon. Highly automated accounts generate a relatively small amount of the content being shared about French politics, though this amount increased over the previous month. Social media users in France shared many links to high quality political news and information with roughly a ratio of 2 links to professionally produced news content for every 1 link to other kinds of political news and information. In comparison to our recently analyzed data from Germany and France, French voters are sharing much better quality information than what many US voters shared, and slightly better quality news and information as German users share.

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 National Science Foundation, "EAGER CNS: Computational Propaganda and the Production / Detection of Bots," BIGDATA-1450193, 2014-16, Philip N. Howard, Principle Investigator and 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 Washington Human Subjects Committee, approval #48103-EG and 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 National Science Foundation or the European Research Council.

REFERENCES

1. Howard, P. N., Kollanyi, B., Bolsover, G., Bradshaw, S. & Neudert, L.-M. Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter? (2017).
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).