

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

COMPROP DATA MEMO 2017.4 / 4 MAY 2017

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

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

Gillian Bolsolver
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 election for three days during the second round of campaigning. For our research on the first round of voting, see Data Memo 2017.3. (1) Content about Macron still dominates Twitter conversation about French politics, though the gap between Macron and Le Pen traffic has narrowed over time. (2) The proportion of traffic originating with highly automated accounts doubled between the first and second round of voting. (3) The ratio of links to professionally produced news content to other political content has gone from about 2 to 1 in the first round of voting to about 1 to 1 in the second round of voting. (4) Compared to our study of similar trends in the US and Germany, we find that Twitter users discussing French politics are sharing higher quality political news and information than many US users were sharing during the US Presidential election and German users were sharing in the German presidential election, but lower quality information than French users were sharing in the first round of voting.

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. However, in our study of bots and junk news in France before the first round of the election, we found that political conversations on Twitter were not as poisoned as the conversations about the 2016 US Election. Twitter users in France shared many links to high quality political news, with roughly a ratio of 2 links to professionally produced content for every 1 link to other kinds of political news and information (a category containing a quarter of junk news, but also a large number of individual blogs and civil society content). Is there any difference in the

kinds of political news and information circulating among voters after the first round of the election?

SAMPLING AND METHOD

The first round of the election took place on April 23 and the second round will take place on May 7. Our dataset contains approximately 960K tweets collected between 27-29 April 2017, using a combination of hashtags associated with the election and the two candidates qualified for the second round, Emmanuel Macron (EM) and Marine Le Pen (FN).

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. 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

962,431 tweets, we selected all of the tweets that contained URLs. Between 27-29 April, Twitter users in France shared 76,063 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. 7,591 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 was carried out by a native French-speaking coder familiar with the media landscape. Once each unique URL was coded, we expanded the coding for our random 10 percent sample.

FINDINGS AND ANALYSIS

Twitter conversation about the French election can be analyzed in terms of the relative use of candidate hashtags, the level of automation, and the kinds of sources for political news and information.

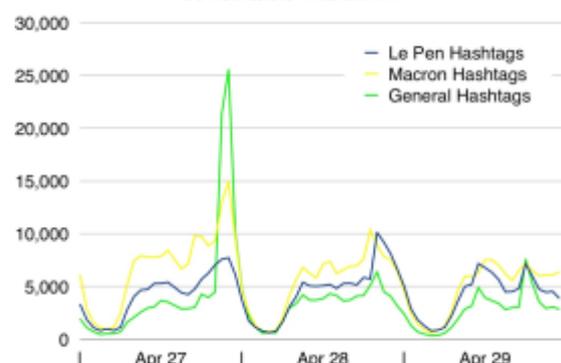
Table 1 and Figure 1 compare the use of candidate hashtags for the sample days in April 2016.

Table 1: Twitter Conversation about the French Presidential Candidates

	N	%
Emmanuel Macron (EM)	412,102	42.3
Marine Le Pen (FN)	310,141	31.8
Neutral	252,805	25.9
Total	975,048	100.0

Source: Authors' calculations from data sampled 04/27-04/29.
 Note: This table is based on the candidate-specific and general hashtags used in the tweets. General hashtags include #présidentielle, #presidentielle, #présidentielles, #presidentielles, #présidentielle2017, #présidentielle2017, #presidentielles2017, #présidentielles2017, #electionpresidentielle, #électionpresidentielle, #electionpresidentielle2017, #electionspresidentielles2017, #elysee2017, #élysée2017, #secondtour, #jevote. Le Pen-specific hashtags: #MarineLePen, #LePen, #MLP2017, #Marine2017, #LePen2017, #AuNomDuPeuple, #FrontNational, #FN, #MarinePresidente, #MarinePrésidente, #JeVoteMarine. Macron-specific hashtags: #Macron, #Macron2017, #EnMarche, #EmmanuelMacron, #MacronPresident, #MacronPrésident, #JeVoteMacron.

Figure 1: Hourly Twitter Conversation about the French Presidential Candidates



Source: Authors' calculations from data sampled 04/27-04/29.
 Note: This figure is based on the candidate-specific and general hashtags used in the tweets.

Hashtags about Emmanuel Macron appeared most often—42.3 percent of the overall traffic relate to him. The peak at the end of the day on April 27 corresponds to his appearance in a TV show. Le Pen appeared in the same show the day before, which is thus not captured in our data. For the most part, traffic about Le Pen is about three-quarter that of Macron’s. Traffic about Le Pen is higher than it was in the study we conducted before the first round: in March, we found that traffic about Le Pen was about half that of Macrons. Overall, traffic about Macron, Le Pen, and the election in general follows the same pattern.

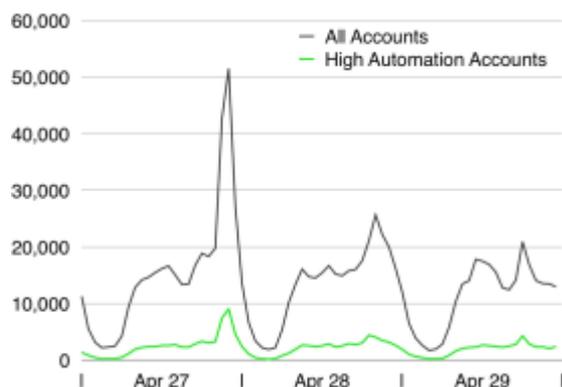
Table 2 and Figure 2 reveal the rhythm of Twitter traffic about the second round of the French Election. We define a high level of automation as accounts that post at least 50 times a day on one of the selected hashtags during the data collection period. This detection methodology fails to capture highly automated accounts that are tweeting with lower frequencies. Both candidates have roughly the same number of highly automated accounts generating traffic about them. In the campaign period for the first round of voting these candidates had over 100 highly automated accounts tweeting with their relevant hashtags, and by the second round over 500 accounts were doing this work for each candidate. However, these highly automated accounts generate different proportions of the candidate’s traffic: 19.5 percent of the Twitter traffic about Macron is driven by highly automated accounts, compared with 14 percent of the Twitter traffic about Le Pen. 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

Table 2: High Frequency Tweeting about the French Election

	N of Tweets	% of Tweets	N of Accounts
Emmanuel Macron (EM)	80,282	19.5	588
Marine Le Pen (FN)	43,564	14	587
General	36,400	14.4	587

Source: Authors’ calculations from data sampled 04/27-04/29.
 Note: This table is based on the candidate-specific and general hashtags used in the tweets.

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



Source: Authors’ calculations from data sampled 04/27-04/29.
 Note: This figure is based on all the hashtags used in the tweets.

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, except for a peak at the time of the TV show, and that it flows in tandem with human generated content during the natural waking hours of human users. On average, 16.4 percent of the traffic (157,638 tweets) about the election is generated by the bots we are able to track. In the sample studied before the first round of the election (a week in March 2017), on average, 7.2 percent of the traffic was generated by bots. In an additional sample collected in February (but not analyzed in depth in a research memo), we found an average of 6.8 percent. By comparison, the level of automation appears to have risen strongly after the first round. However, it must be noted that we changed the set of hashtags after the first round to include new general hashtags and hashtags relevant to the two selected candidates only. Furthermore, the traffic we captured with this new set of hashtags has strongly increased compared to our previous study: we had roughly 120K tweets per day in our March dataset, compared to 320K tweets per day in this dataset.

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.

- 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.
- Humor and Entertainment. Content that involves political jokes, sketch comedy, political art or lifestyle- or entertainment-focused coverage.
- Religion. Links to political news and information with distinctly religious themes and faith-based editorializing presented as political news or information.
- Russia. This content was produced by known Russian sources of political news and information.
- 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
 - 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.
 - 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
 - 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.
 - 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 the second round of the French election comes from professional news organizations. However, 29.4 percent of the content shared involves other kinds of political news and information. The largest proportion of that content is citizen-generated content (such as personal and organizational blogs) and other political content (such as portals and aggregators). Yet together, junk news, Russian content, and religious content account for a third of the category. Overall, junk news, Russian content, and religious content account for 10 percent of all links shared over Twitter (6 percent for junk news alone). By comparison, this proportion was 7 percent in the similar study we conducted before the first round (4 percent for junk news alone). In other words, there was a 50 percent increase since our last study. There are other ways to see that French voters appear to be sharing political news and information of a lesser quality between the two rounds compared to before the first round of the election. Before the first round, we found that almost 47 percent of URLs

shared were professional news content and 16 percent were professional political content produced by government, political parties, and experts. In this study, we find that these proportions have decreased by 7 and 4 percentage points respectively. However, French voters still share political news and information of better quality than US voters: in our study of Michigan, we found a 1:1 ratio between professional news and junk news alone.

Table 3: French Political News and Information On Twitter

<i>Type of Source</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Professional News Content				
Major News Brands	2,752	90.2		
Minor News Brands	299	9.8		
Subtotal	3,051	100.0	3,051	40.2
Professional Political Content				
Political Party or Candidate	794	88.3		
Experts	56	6.2		
Government	49	5.5		
Subtotal	899	100.0	899	11.8
Other Political News and Information				
Other Political	746	33.4		
Citizen or Civil Society	639	28.6		
Junk News	468	21.0		
Russia	243	10.9		
Humor or Entertainment	91	4.1		
Religion	35	1.6		
Political Merchandise	5	0.2		
WikiLeaks	4	0.2		
Subtotal	2,231	100.0	2,231	29.4
Other				
Social Media Platform	910	82.1		
Other Non-Political	198	17.9		
Subtotal	1,108	100	1,108	14.6
Inaccessible				
Language	269	89.1		
No Longer Available	33	10.9		
Subtotal	302	100.0	302	4.0
Total			7,591	100.0

Source: Authors' calculations from data sampled 04/27-04/29.

CONCLUSIONS

The internet has long been a means of manipulating public opinion.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 different countries and at different times. French social media users appear to share more junk news between the two rounds of the French election than they did before the

first round. Still, conversations about the French election over social media are not as poisoned as conversations about the 2016 US election, where we found as much junk news as professional news shared over Twitter by Michigan voters. Even if French voters share less links to high quality news between the two rounds than they did before the first round, substantive differences between the qualities of political conversations in France and in the US remain evident. 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 government agencies, traditional political parties, or other experts. In France, between the two rounds, 41.7 percent of all links led to professional news and 8.8 percent to content produced by government agencies, political parties, and experts. French voters discussing the second round of the presidential election over social media tend to use more high quality information sources than US voters discussing the 2016 US election, but slightly less than when they discussed the first round.

Content about Macron still tends to dominate the traffic on Twitter between the two rounds, but content about Le Pen increased compared to before the first round. Even if comparisons are made difficult by the fact that we changed our set of hashtags to adjust to the second round, it appears that traffic about the election on Twitter strongly increased. A larger part of this traffic comes from bots: highly automated accounts appear to generate twice as much traffic as they did in the months before the first round. Social media users in France still share better quality political news and information than what many US voters shared, but they share less quality information when discussing the second round than they did when discussing the first round. Before the first round, we found that Twitter users in France shared links with a ratio of approximately 2 links to professionally produced news content for every 1 link to other kinds of political news and information. In comparison, between the two rounds, we find a ratio of approximately 1 link to professional news for 1 link to other kinds of political news and information.

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).