Junk News and Bots during the 2017 UK General Election: What Are UK Voters Sharing Over Twitter?

COMPROP DATA MEMO 2017.5 / 31 MAY 2017

John D. Gallacher
Oxford University
john.gallacher@cybersecurity.ox.ac.uk
@john_gallacher1

Monica Kaminska
Oxford University
monica.kaminska@cybersecurity.ox.ac.uk
@monica_kaminska

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

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

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 2017 UK General Election for a week in May 2017. (1) Content about the Labour Party tended to dominate traffic on Twitter. (2) Automated accounts generated a relatively small amount of content about UK politics, and while this automation was spread fairly equally across parties, highly automated accounts associated with the Labour Party were more active in generating traffic. (3) Social media users in the UK shared four links to professional news and information for every one link to junk news. (4) In comparison to our study of similar trends in the US, Germany and France, we find that UK users shared better quality information than that which many US users shared, but worse quality news and information than German and French users shared.

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, 'robotic' tasks. They can be used to perform legitimate tasks like delivering news and information—real news as well as junk—or undertake malicious activities like spamming, harassment and hate speech. Whatever their uses, highly automated social media accounts are able to rapidly deploy messages, replicate themselves, and pass as human users. They are 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 on a global scale. During the 2016 UK Brexit referendum it was found that political bots played a small but strategic role shaping Twitter conversations.[6] The family of hashtags associated with the argument for leaving the EU dominated, while less than 1% of sampled accounts generated almost a third of all the messages.

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.[7] 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.[8] 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.[9] Both state and non-state political actors can deliberately manipulate and amplify non-factual information online.

Junk news websites deliberately publish misleading, deceptive or incorrect information purporting to be real news about politics, economics or culture.[10] These sites often rely on social media to attract web traffic and drive engagement. Both junk 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 UK voters? 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?

SAMPLING AND METHOD

Our dataset contains approximately 1,363,000 tweets collected between the 1st and the 7th of May 2017, using hashtags associated with the primary political parties in the UK, the major candidates, and the election itself.

Twitter provides free access to a sample of 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 at any time.[11] 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. We used two sets of hashtags. The first set was used to collect URLs that people were sharing as part of the wider election conversation (see Table 3). A subset of these hashtags (see Table 1) was then selected to collect information about how much Twitter conversation each party was generating and how much of this was automated.

Parliamentary and multi-party systems tend to have more variety of hashtags related to particular candidates and important political issues. Thus, our sampling strategy may have missed 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 was done in the R software environment developed for statistical computing.

Selecting tweets based on hashtags has the advantage of capturing the content most likely to be relevant to the currently studied political event. The Streaming API yields (1) tweets which contain the selected hashtags; (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 a hashtag; (3) retweets that contain a message's original text, wherein the 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 the 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 1,362,666 tweets, we selected all of the tweets that contained URLs. Between the 1st and the 7th of May, Twitter users in the UK shared 184,580 links on the platform. URLs that pointed towards another tweet were removed from our sample, as most of these tweets are 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 analysed. We then generated a random 10% sample of the dataset using a Python script, which contained 18,457 URLs. We removed duplicate URLs from our sample to classify each URL according to our classification system. The classification of each URL was carried out by a team of six coders fluent in the English language and familiar with the media landscape. They worked together over a period of two days, and to ensure consistency across coders a training period was carried out, followed by a short test of ground-truth URLs which all coders were required to pass. Once each unique URL was coded, we expanded the coding to the duplicate URLs to complete the coding for our random 10% sample.

The grounded typology of news platforms and content types that was used is as follows:

- 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.
  - 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.

FINDINGS AND ANALYSIS

Twitter conversation about UK politics can be analyzed in terms of the relative use of party hashtags and 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 party specific hashtags for the sample week in May 2017. Hashtags about the Labour Party appeared most often, representing 39.7% of the party-specific tweets during the week as a whole. The Conservative Party generated the second highest proportion of the conversation at 26%, however Figure 1 reveals that at times the Conservative Party conversation was higher than that of the Labour Party. At 19%, the Scottish National Party generated a disproportionally high percentage of the conversation, especially given the size of the party. UKIP and the Liberal Democrats generated 9.6% and 5.7% of the traffic respectively.

Table 2 and Figure 2 reveal the rhythm of Twitter traffic about the UK General 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 political party, it appears that the Conservative Party and the Labour Party have a higher number of highly automated accounts generating traffic about them when compared to the other three parties. These highly automated accounts generate similar proportions of traffic across the Labour Party, the Conservative Party, the Liberal Democrats and UKIP, however highly automated accounts associated with the Labour Party were more productive in generating a higher overall number of tweets for a similar number of automated accounts. A higher proportion of tweets about the SNP, 15.6%, was generated by highly automated accounts in comparison to other parties.

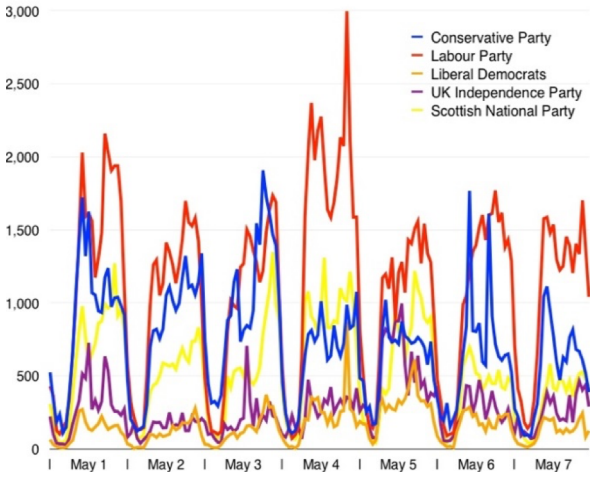
We cannot know who manages these accounts, and we do not analyze the content or emotional valence of particular tweets. Hence, this

Table 1: Twitter Conversation about the UK Election

	Number of tweets	%
Labour Party	185,602	39.7
Conservative Party	121,387	26.0
Scottish National Party (SNP)	88,673	19.0
UK Independence Party (UKIP)	45,083	9.6
Liberal Democrats	26,827	5.7
Total	2,890,867	100.0

Source: Authors' calculations from data sampled 1-7 May 2017.
Note: Conservative hashtags include #TheresaMay, #Tories, #Tory, #AskTheresaMay, #Conservatives, #VoteTory, #StrongAndStable; Labour hashtags include #VoteLabour, #Labour, #JC4PM, #Corbyn, #LabourDoorstep, #JeremyCorbyn, #Corbyn4pm, #JezzWeCan, #VoteCorbyn; Liberal Democrat hashtags include #LibDems, #LibDemFightBack, #LibDem, #TimFarron, #UniteforEurope, #VoteLibDem, #LibDemSurge; UKIP hashtags include #UKIP, #Farage, #VoteUKIP, #Nuttall; SNP hashtags include #ScotRef, #IndyRef2, #VoteSNP, #SNP, #NicolaSturgeon.

Figure 1: Hourly Twitter Conversation about the Major Parties in the UK Election



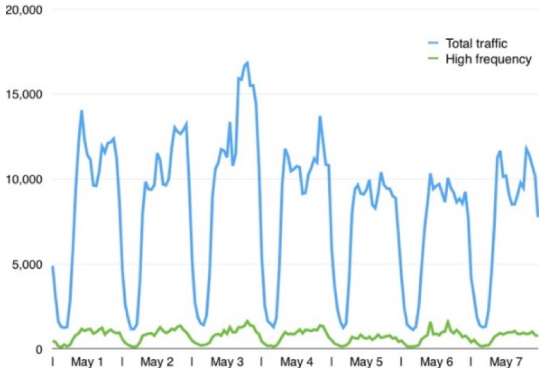
Source: Authors' calculations from data sampled 1-7 May 2017.
Note: This figure is based on the party-specific hashtags used in the tweets.

Table 2: High Frequency Tweeting about the UK Election

	N of Tweets	% of Total	N of Accounts
Labour Party	21,661	11.7	179
Conservative Party	13,409	11.1	182
Scottish National Party (SNP)	13,819	15.6	147
UK Independence Party (UKIP)	5,167	11.5	163
Liberal Democrats	3,399	12.7	145

Source: Authors' calculations from data sampled 1-7 May 2017.

Figure 2: High Frequency Tweeting on the UK Election, Hourly



Source: Authors' calculations from data sampled 1-07 May 2017.

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 UK 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 12.3% of traffic about UK politics is generated by highly automated accounts that we are able to track.

To understand what kinds of political news and information UK voters are sharing, we then analyzed the links included in the tweets that contained our selected hashtags about the UK election. Table 3 explains the distribution of content shared by UK Twitter users and reveals that the largest proportion of content being shared by Twitter users interested in UK politics comes from professional news organizations, which accounts for 43.3% of the total content shared. Junk news accounts for over a third of other political news and information and accounts for 10.2% of the total content shared.

Within the professional news content that was shared, daily newspaper The Guardian was the most popular, with 15.2% of professional news coming from this source. This was followed by the BBC with 10.4% of links directing to BBC content. A high percentage of other political content that was shared comes from citizen- generated sources such as personal blogs or civil society organizations. The number of links to such sources were of a comparable level to that of junk news.

Russian sources did not feature prominently in the sample, and no content was shared that could be attributed to WikiLeaks in contrast our project's previous memos on the US and French elections.[1,12,13]

Incidentally, the number of links to other non-political content was relatively high and included links to non-political company websites, large online retail organizations, and spam content.

Having performed this analysis over four major elections in the past twelve months, we can now compare the consumption of professional news across several countries. Table 4 shows the levels of automation and junk news shared on Twitter across the major global elections that have occurred so far in 2016-2017. To better aid comparison across countries, these figures display the percentages of junk news once content that was allocated to the category 'Other' and inaccessible content was removed, leaving only relevant content.

UK social media users share a higher percentage of junk news content than social media users who are actively discussing German politics and French politics during election periods. The proportion of relevant content shared on UK social media identified as junk news was 12.6%, compared to 12.5% in Germany and 5.1% and 7.6% respectively in the two election rounds in France. Yet we also found that UK users were not sharing as much junk news in their political conversations as US users in the lead up to the 2016 elections, where the level of junk news shared was significantly higher. In the days leading up to the US election, we did a close study of junk news consumption among Michigan voters and found users were sharing as much junk news as professional news content at around 33% of total content each.

Substantive differences between the qualities of political conversations are evident in other ways. In the US sample, 33.5% of relevant links being shared led to professional news content. In Germany this was 55.3%, and in France this was between 49.4% and 57% of relevant links across both election

Table 3: UK Political News and Information On Twitter

Type of Source	*N*	*%*	*N*	*%*
Professional News and Information				
Major News Brands	5,586	69.9		
Minor News Brands	2,403	30.1		
Subtotal	7,989	100.0	7,989	43.3
Professional Political Content				
Political Party or Candidate	805	48.6		
Government	625	37.8		
Experts	225	13.6		
Subtotal	1,655	100.0	1,655	9.0
Other Political Content				
Citizen or Civil Society	2,102	39.6		
Junk News	1,882	35.4		
Other Political	887	16.7		
Russia	151	2.8		
Humor or Entertainment	210	4.0		
Religion	65	1.2		
Political Merchandise	14	0.3		
Subtotal	5,311	100	5,311	28.8
Other				
Social Media Platform	1,046	37.1		
Other Non-Political	1,774	62.9		
Subtotal	2,820	100	2,820	15.3
Inaccessible				
Language	425	62.3		
No Longer Available	257	37.7		
Subtotal	682	100	682	3.7
Total			18,457	100

Source: Authors' calculations from data sampled 1-7 May 2017. Note hashtags include: #generalelection, #ge2017, #brexit, #ge17, #generalelection2017, #remain, #election2017, #stopbrexit, #theresamay, #toriesout, #toryelectionfraud, #tories, #tory, #asktheresamay, #conservatives, #votetory, #strongandstable, #publicduty, #votenhs, #makejunetheendofmay, #electionfraud, #votelabour, #labour, #jc4pm, #corbyn, #labourdoorstep, #jeremycorbyn, #corbyn4pm, #jezzwecan, #votecorbyn, #libdems, #libdemfightback, #libdem, #timfarron, #uniteforeurope, #votelibdem, #libdemsurge, #ukip, #farage, #voteukip, #nuttall, #scotref, #indyref2, #votesnp, #snp, #scotland, #nicolasturgeon.

Table 4: Automation and Junk News in Major Elections, 2016-2017

Country	Percent of relevant content from professional news sources	Percent of relevant content from parties, government agencies, or experts	Percent of content from automated sources	Percent of relevant content that is "Junk News"	Percent of relevant content from Russian news sites	Ratio of links to professionally produced news to other political content	Ratio of links to professionally produced news to junk news
USA – Michigan. Sample, 1-11 November 2016, 22m tweets.	33.5	4.4	-	33.8	1.1	0.5:1	1.0:1.0
Germany. Sample, before voting, 1-13 February 2017, 121K tweets.	55.3	16.8	5.7	12.5	3.3	2.0:1	4.4:1
France I. Sample, before Round 1 voting, 13-19 March 2017, 842K tweets.	57.0	19.2	7.2	5.1	3.0	2.4:1	11.2:1
France II. Sample, between Round 1 and 2, 27-29 April 2017, 960K tweets.	49.4	15.4	16.4	7.6	3.9	1.4:1	6.5:1
United Kingdom. Sample, soon after election announced, 1-7 May 2017, 1.4M tweets.	53.4	11.1	12.3	12.6	1.0	1.5:1	4.2:1

Note: 'Relevant content' is calculated after other non-political content, spam, irrelevant social media, language and inaccessible content have been removed.

rounds. Similarly, in the current UK-based study we show that 53.4% of relevant links being shared led to professional news content. Individuals discussing politics over social media in the European countries sampled tend to share more high quality information sources than US users.

CONCLUSIONS

The Internet has long been used both for political activism and social control.[14] 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 UK politics.

Content about the Labour Party tended to feature prominently among the election traffic on Twitter. The level of automation was roughly equal across the Conservative Party, Labour Party, Liberal Democrats and UKIP, however highly automated accounts tweeting about the Labour Party were more productive in spreading content. Overall though, automated accounts generate a relatively small amount, 12.3%, of the total content being shared about the UK election. When evaluating the quality of information shared, we found that social media users in the UK shared roughly 1.5 links to professionally produced news sources to every 1 link to other kinds of political news and information. In comparison to our recently analyzed data from Germany, France, and the US, UK users were sharing better quality information than many US users, but lower quality news and information than French and German users.

ABOUT THE PROJECT

The Project on Computational Propaganda (www.comprop.oii.ox.ac.uk) 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, 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 and the Engineering and Physical Sciences Research Council (EPSRC). 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, the European Research Council or the Engineering and Physical Sciences Research Council or the University of Oxford.

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. Howard, P. N. & Kollanyi, B. Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum. ArXiv160606356 Phys. (2016).
7. World Economic Forum. 10. The Rapid Spread of Misinformation Online. *Outlook on the Global Agenda 2014* (2014).
8. Vicario, M. D. *et al.* The Spreading of Misinformation Online. *Proc. Natl. Acad. Sci.* 113, 554–559 (2016).
9. 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).
10. Howard, P. N. Digitizing the social contract: Producing American political culture in the age of new media. *Commun. Rev.* 6, 213–245 (2003).
11. 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).
12. Howard, P. N., Bradshaw, S., Kollanyi, B., Desigaud, C., & Bolsover, G. Junk News and Bots during the French Presidental Election: What Are French Voters Sharing Over Twitter? (2017).
13. Desigaud, C., Howard, P. N., Bradshaw, S., Kollanyi, B., & Bolsover, G. Junk News and Bots during the French Presidental Election: What Are French Voters Sharing Over Twitter In Round Two? (2017).
14. Howard, P. N. *Pax Technica: How the Internet of Things May Set Us Free or Lock Us Up*. (Yale, 2015).