

Russian Involvement and Junk News during Brexit

[COMPROP](#) DATA MEMO 2017.10 / December 19th 2017

Vidya Narayanan
Oxford Internet Institute
Oxford University
vidya.narayanan@oii.ox.ac.uk
[@vidunarayanan](https://twitter.com/vidunarayanan)

Philip N. Howard
Oxford Internet Institute
Oxford University
philip.howard@oii.ox.ac.uk
[@pnhoward](https://twitter.com/pnhoward)

Bence Kollanyi
Oxford Internet Institute
Oxford University
bence.kollanyi@oii.ox.ac.uk
[@bencekollanyi](https://twitter.com/bencekollanyi)

Mona Elswah
Oxford Internet Institute
Oxford University
mona.elswah@oii.ox.ac.uk
[@monaelswah](https://twitter.com/monaelswah)

ABSTRACT

Several prominent UK politicians are concerned about the role of Russian troll and bot accounts in public conversation over social media during the Brexit debate in 2016. Looking at our archive of Twitter conversation we find that (1) Russian Twitter accounts shared to the public, contributed relatively little to the overall Brexit conversation, (2) Russian news content was not widely shared among Twitter users, and (3) only a tiny portion of the YouTube content was of a clear Russian origin. It is possible that there are other accounts, not yet discovered, that influenced Brexit conversations, and that there was other activity on platforms like Instagram, or Facebook. We conclude with some observations about the impact of strategically disseminated polarizing information on public life.

COMPUTATIONAL PROPAGANDA AND THE BREXIT DEBATE

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’ work. 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.

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 in 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.

In recent weeks, reports claiming varying levels of Russian activity in the Brexit referendum, have appeared in the media. The main research in Russian interference during the Brexit referendum was carried out at the following universities: (1) Swansea University, (2) City University, London, (3) University of Edinburgh and (4) University of Oxford.

The research group based at Swansea collected Brexit data on Twitter, between the months of May and August 2016 and showed that public opinion on Twitter could predict the outcome of the referendum.¹⁴ To investigate the spread of public opinion before, during and after the referendum. They used the Twitter streaming API with #Brexit, to collect a total of 28M tweets, between 24th May and 17th August 2017. The data sample contained details such as, user names, date of account creation, date of tweeting, number of retweets and the number of followers. Then they examined both pro-leave

tweets and pro-remain tweets using methods outlined in previous research work and extracted the relevant hashtags.¹³

They classified 20% of the accounts as bots using three criteria viz., abnormal time of tweeting (00:00 hrs - 06:00 hrs, BST), abnormal number of tweets per day and tweets from platforms. Their research indicates that bots on Twitter can influence public opinion and that the outcomes of major events that involve public participation, can be predicted by studying bot interactions on Twitter and that bots influenced the outcome of the Brexit referendum. Using the list of bot accounts that Twitter has made public, and linking these to the data sample they collected during Brexit, media reports claim that they have discovered 150,000 bot accounts linked to Russia that have been active during Brexit. However, they are yet to produce a formal publication, where these claims linking the accounts to the Russian based Internet Research Agency, have been verified.

The researchers at City University also used the Twitter streaming API, to search for Brexit related tweets, between April 2016 and August 2016.¹⁵ They investigated bot activity to establish the following: (1) dissemination of hyper-partisan and polarizing content, (2) to characterize the lifecycle of a bot - with a period of high posting activity followed by a sharp drop in activity levels, (3) to determine the influence bots had on Brexit discussions, (4) to determine if bots caused faster cascades than human users, looking at the impact through spread, reach and intensity, (5) to distinguish bots from human users and other bots in the botnet, by measuring if the botnet was located in a network of human users or if was restricted to only to clusters of botnets.

The study analyzed 10M tweets, extracted from the API, between June 10th, 2016 and July 10th, 2016, that referenced the referendum using relevant hashtags. They used, a number of metrics like the presence or absence of geographical metadata, account creation date, followers to following ratio and activity levels, to distinguish human users from bots. According to these researchers, some positive predictors of bot activity include tweets to user,

uncommon words in user name, use of web interface to post content, retweet reciprocity and account creation date.

Out of a total of 794,949 users, only 30,122, (37%) were located in the UK, 40,031 users have either been deleted, changed to private, blocked, deactivated or changed their usernames since the referendum. Further, 17% of tweets had the keyword 'remain' while 31% of the messages contained the word 'leave'. Also, bots were eight times more likely to tweet leave slogans than other Twitter users and 63% of URLs in bot tweets do not exist any longer. Media reports, suggest that the researchers have identified 13,493 bot accounts with Russian links.

Media reports also claim that, the research group working in the area of politics and public policy at The University of Edinburgh, have identified 419 bot accounts with links to Russian agencies and Yin Lu a research student working at the Oxford University, has identified 54 Russian bot accounts.¹⁶

JUNK NEWS AND AUTOMATION DURING BREXIT

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. In this paper we seek to (1) identify highly automated accounts that were either linked to Russia or tweeting pro-Russian content, (2) classify the type of content that was shared by these accounts and finally (3) analyze YouTube videos shared by a sample of Twitter users who were active during the referendum, and classify the type of political news stories shared into high-quality, professional news, extremist, sensationalist, conspiratorial content, masked commentary, fake, and other forms of junk news and news stories linked to Russian sources.

SAMPLING AND METHOD

This data set contains 5,811,102 Tweets collected between the 6-12th June and again 17-23 June 2016, using a

combination of pro-leave, pro-remain and neutral hashtags to collect the data. This sampling strategy yielded tweets from 1,112,403 distinct Twitter user accounts. Since our purpose is to discern how bots were being used to amplify political communication on this important policy question, we did some basic descriptive analysis to understand the rhythm of social media activity on this topic. Given the limits that Twitter places on researchers, it is impossible to report the total number of bots engaged in the StrongerIn-Brexit debate.¹⁰

These tweets and associated data were collected from Twitter's public streaming API at the time of the election, not retroactively with the Search 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.⁶ Tweets were selected based on a list of hashtags associated with the UK referendum, and tweets were collected from the API that (1) contained at least one of the relevant hashtags; (2) contained the hashtag in the text of a link, such as a news article, shared in that tweet; (3) were retweets of a message that contained the hashtag in the original message; or (4) quoted tweets in which the hashtag was included but in which the original text was not included and Twitter used a URL to refer to the original tweet.

Unfortunately, it is difficult to reconstruct social networks of Twitter users, years after their activity. Accounts that were tweeting about Brexit using other hashtags are not in this data set. There may be Russian origin accounts in our dataset that we have misclassified because the evidence of this association was in previous Twitter activity or a network tie that was not captured in our collection efforts at the time.

To evaluate different sources being shared over social media, we determined the source of each of the URLs in the dataset. Overall, 1,429,182 of the 5,811,102 tweets contained a URL. In order to reveal the URLs that were shortened through one of the major URL shortening services, such as bit.ly, goo.gl, fb.me, and tiny.url, we first took a random 10 percent sample of all the tweets with a shared link. Then, we used a script to visit every shortened URL and save the address of the original web page. Based on a dictionary of domains that belong to an easy to identify source, such as various media sources, social media platforms, and political parties, a large number of URLs were auto-coded. Effectively this typology is built on the successful cataloguing of 87,169 links out of a total of 142,918 links in our 10 percent random sample. This typology that has emerged over our study of elections in five democracies. The grounded typology of news platforms and content types that was used is as follows:

- Professional News and Information
 - 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 displaying 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 Content
 - Junk News. This content includes various forms of propaganda and ideologically extreme, hyper-partisan, or conspiratorial political news and information. 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 it 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 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 Brexit-related hashtags. Spam is also included in this category.

- Inaccessible
 - 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.
 - No Longer Available. These links were shared, 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 URL could be attributed to another source, then it is classified according this information.

FINDINGS

First, we examined the Brexit related Twitter traffic from a list of automated accounts linked to Russia. For this, we took the list of accounts that Twitter recently identified as

Table 1: Political News and Information on Twitter during the Brexit debate

<i>Type of Source</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Professional News and Information				
Major News Brands	48,132	86.4		
Minor News Brands	7,587	13.6		
Subtotal	55,719	100.0	55,719	63.9
Professional Political Content				
Government	4,160	11.0		
Political Party	844	15.0		
Experts	617	11.0		
Subtotal	5,621	100.0	5,621	6.5
Other Political Content				
Citizen or Civil Society	2,756	33.3		
Junk News	3,650	44.0		
Online Portal	720	8.7		
Humor or Entertainment	631	7.6		
Russia	511	6.2		
Religion	19	0.2		
Political Merchandise	0	0.0		
Subtotal	8,287	100.0	8,287	9.5
Relevant Content Subtotal			69,627	79.9
Other				
Social Media Platform	16,843	96.0		
Other Non-Political	699	4.0		
Subtotal	17,542	100.0	17,542	20.1
Total			87,169	100

Source: Authors' calculations from data sampled 6–12 June and 17–23 June 2016.

Note: Hashtags include #brexit, #voteleave, #leaveeu, #takecontrol, #betteroffout, #voteout, #believe, #brexithemovie, #euistheproblem, #brexitbustour, #strongerin, #remain, #voteremain, #votein, #bremain, #labourin, #votestay, #intgether, #labourinforbritain, #greenerin, #euref, #eureferendum, #inorout, #eudebate, #june23.

Table 2: Political News and Information in YouTube shared by Twitter users

Type of Source	N	%	N	%
Professional News and Information				
Major News Brands	133	56.8		
Minor News Brands	101	43.2		
Subtotal	234	100	234	7.9
Professional Political Content				
Experts	118	57.3		
Political Party	86	41.7		
Government	2	1.0		
Subtotal	206	100	206	6.9
Other Political Content				
Citizen or Civil Society	668	51.7		
Junk News	359	28.0		
Humor or Entertainment	156	12.0		
Online Portal	48	3.7		
Russia	41	3.2		
Religion	17	1.3		
WikiLeaks	2	0.2		
Political Merchandise	0	0.0		
Subtotal	1291	100	1291	43.8
Relevant Content Subtotal			1554	
Other				
Other Non-Political	752	63.1		
Not Available	394	33.1		
Inaccessible Language	43	3.6		
Lifestyle	25	2.0		
Shopping	2	0.2		
Subtotal	1216	100	1216	41.3
Total				
	2947	100	2947	100

Source: Authors' calculations from data sampled 6–12 June and 17–23 June 2016.

Note: Hashtags as in Table 1.

being managed by the Russian based Internet Research Agency (2,752 accounts). Then we supplemented the list with additional accounts that a) were identified by other analysts (94 accounts) and b) were reported in major news media (51 accounts). We followed a systematic approach to search for media reports of bots that were controlled by Russian interests, using the following search strings, “Russia bots”, “Brexit bots”, “Brexit Russia”, “Brexit Twitter”, “Russia Twitter”, “Russian Twitter bots”, using both Google News and the LexisNexis database, thus ensuring that the search was comprehensive. Altogether only 105 of these accounts produced any tweets or retweets about Brexit, and they generated only 15,964 tweets in our sample, which represents less than 0.3% of the total traffic. We also found that overall, only 3% of the Russian accounts initially “outed” by Twitter generated any traffic during the Brexit debate.

Second, looking at the news content shared during these Brexit conversations on Twitter, it appears that the proportion of content from known Russian sources like Russia Today or Sputnik got little traction over Twitter. Only 511 of the 87,169 links shared were to these sources, representing under 0.6% of the data. Perhaps what is more concerning than links to Russian news sources, is the volume of junk news shared by the Twitter users. The junk news content was more than 7 times the volume of Russian sourced news.

The distribution of content shared by UK Twitter users is shown in Table 1. The largest proportion of content shared by Twitter users interested in UK politics comes from professional news organizations, which accounts for 63.9% of the URLs shared on Twitter in our sample. Within that, the Telegraph was most popular, with 8.1% of professional news coming from this source. This was followed by the BBC with 6.2% of links directing to its website.

Junk news accounts for almost half of other political news and information. A high percentage of other political content that was shared comes from citizen-generated sources like personal blogs or civil society organizations. The number of links to such sources was higher than the number of links to junk news. Russian sources did not feature prominently in the sample, corresponding to our earlier UK election study and no content was shared that could be attributed to WikiLeaks. This was in contrast to our project’s previous memos on the US and French elections.

Third, we took a dedicated look at the YouTube videos shared by twitter users during the referendum. For this, we selected a random sample of 10% of all Twitter users’ data, collected using the Brexit related hashtags. We coded all the YouTube videos on the basis of the typology and the distribution of video content across categories is shown in Table 2. Using our standard dictionary, we labelled over 11.0% of the content as junk news and only 8.0% of the content as sourced from professional news channels. Over 22.7% of the content was created by citizens or civic society groups and nearly 5.3% of videos were classified as political humor or satire and finally only 1.4% of the videos were linked to Russian news sources.

CONCLUSIONS

From previous studies, it seems clear that highly automated accounts were active on Twitter, in the period leading up to the Brexit referendum. However, in our investigations of Twitter conversations associated with highly automated accounts we found little evidence of links to Russian sources. On examining a sample of YouTube links tweeted by Twitter users during the referendum, we again found little evidence of Russian content. A matter of concern however is the large number of accounts both human and automated, that shared polarizing and provocative content over the social media platform in days leading up to the referendum.

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. The project gratefully thanks the Ford Foundation for their support. 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 University of Oxford or the European Research Council.

REFERENCES

1. Gallacher, J., Kaminska, M., Kollanyi, B., Yasseri, T. & Howard, P. N. Social Media and News Sources during the 2017 UK General Election. (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. Kumpel, A. S., Karnowski, V. & Keyling, T. News Sharing in Social Media: A Review of Current
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., Bolsover, G., Kollanyi, B., Bradshaw, S. & Neudert, L.-M. Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter? (2017).
12. Kollanyi, B., Howard, P. N. & Woolley, S. C. *Bots and Automation over Twitter during the Third U.S. Presidential Debate*. 4 (Project on Computational Propaganda, 2016).
13. Howard, P., Kollanyi, B. & Woolley, S. Bots and Automation over Twitter during the U.S. Election. *Oxf. UK Proj. Comput. Propag.* (2016).
14. Hern, A. Russian troll factories: researchers damn Twitter’s refusal to share data. *The Guardian* (2017).
15. Bastos, M. T. & Mercea, D. The Brexit Botnet and User-Generated Hyperpartisan News. *Soc. Sci. Comput. Rev.* 0894439317734157 (2017). doi:10.1177/0894439317734157
16. Russian Twitter accounts tweeted 3,468 times about Brexit. *The Independent* (2017). Available at: <http://www.independent.co.uk/news/uk/politics/brexit-latest-russian-twitter-accounts-eu-independence-referendum-tweets-influence-result-a8055746.html>.