

Junk News and Bots during the German Federal Presidency Election: What Were German Voters Sharing Over Twitter?

[COMPROP](#) DATA MEMO 2017.2 / 27 March 2017

Lisa-Maria Neudert
Oxford University
lisa-maria.neudert@oii.ox.ac.uk
[@lmneudert](https://twitter.com/lmneudert)

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

Philip N. Howard
Oxford University
philip.howard@oii.ox.ac.uk
[@pnhoward](https://twitter.com/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 data on bot activity and junk news using a set of hashtags related to the German Federal Presidency Elections in February 2017. We find that (1) traffic about the far-right Alternative für Deutschland and their candidate Albrecht Glaser accounted for a surprisingly large portion of Twitter activity given their share of voter support. (2) Overall, the impact of political bots was minor, with highly automated accounts generating a small fraction of the Twitter traffic about the election. (3) Social media users in Germany shared many links to political news and information, and the ratio of professional news to junk news shared by German Twitter users was 4 to 1.

AUTOMATION AND SOCIAL MEDIA

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 What kinds of political news and information were circulating over social media among voters in Germany during the recent Presidential election? How much of it was extremist, sensationalist, conspiratorial, masked commentary, fake, or some other form of junk news?

Junk news, widely distributed over social media platforms, can in many cases be considered to be a form of computational propaganda. Social media platforms have served significant volumes of fake, sensational, and other forms of junk news at sensitive moments in public life, though most platforms reveal little about how much of this content there is or what its impact on users may be. The World Economic Forum recently identified the rapid spread of misinformation online as among the top 10 perils to society.⁶ Prior research has found

that social media favors sensationalist content, regardless of whether the content has been fact checked or is from a reliable source. 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.^{7,8} Both state and non-state political actors deliberately manipulate and amplify non-factual information online.

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

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

How were highly automated accounts used around Federal Presidency Election Day 2017 in Germany?

COMPUTATIONAL PROPAGANDA IN GERMANY

There have been some dramatic examples of computational propaganda in Germany in recent months. Angela Merkel was bombarded with bot-generated hate speech messages after the Berlin Christmas market attacks in December 2016. Automated right-wing accounts on Twitter propagate

xenophobic messages in relation to the German refugee debate. Presumed bot networks supporting the Alternative für Deutschland (AfD) have been discovered on Facebook.^{10,11}

In November 2016, Chancellor Angela Merkel warned the German Bundestag about the potential influence of social bots and digital misinformation on the formation of public opinion and their potential to tamper with it. All of the major German parties including the Sozialdemokratische Partei Deutschlands (SPD), Christlich Demokratische Union / Christlich-Soziale Union (CDU / CSU), Bündnis90 / Die Grünen and Die Linke have publically stated that they would refrain from using social bots in elections and strongly condemn their employment. The right-wing AfD, in contrast, stated that they would “consider the use of social bots for elections”. However, the party distanced itself from this statement later.^{12,13}

Several political leaders are proposing greater public policy oversight of the platforms and algorithms that make computational propaganda possible. Bündnis90 / Die Grünen demands mandatory labelling for bots on social media and the governing party CDU / CSU has proposed a counterstatement obligation to fake news on social media. Three German states have re-proposed a legislative initiative on digital trespassing that would prohibit the use of social bots online altogether when implemented. In March 2017, Germany’s Justice Minister Heiko Maas proposed a law that would impose heavy fines on social network sites if they fail to take down illegal hate speech and junk news content. Experts replied that such regulations might be overbearing and negatively affect freedom of expression.¹⁴ But just how active are bots in German public life, and how many fake news stories are Germans consuming and sharing?

SAMPLING AND METHOD

To understand the phenomena, we constructed a dataset containing 121,582 tweets collected February 11–13, 2017 using a combination of hashtags associated with the Presidential election, the candidates, and the parties they represent. Since our purpose is to discern how bots are being used to amplify political communication, we did specific analysis of hashtags used in this dataset.

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.¹⁵ In order to get the most complete and relevant data set, the tweets were collected by following particular hashtags identified by the team as being actively used during the election. A few additional hashtags were added on the day of the election as they rose to prominence. The

programming of the data collection and most of the analysis were done by using Python and R.

Selecting tweets on the basis of 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. Tweets from people who did use these hashtags, but were tweeting about something else, are captured in this sample. The content about the Presidential elections almost exclusively used neutral hashtags. Prevalent mobilizing hashtags were only identified in relation to the AfD, but the mere usage of hashtags there reveals little about the political affinity of the user.

FINDINGS AND ANALYSIS

Of the total of tweets captured in this sample, some 17,453 tweets included links to external content. A random sample of 10 percent of these tweets containing a URL was drawn and analyzed. The linked webpages from the subset were identified and classified into content categories distinguishing political news and information. The set of all 17,453 URLs was then screened for other instances of these categorized sources. Roughly 94 percent of URLs were identified using this approach. The sample contains 14,852 tweets on political news and information.

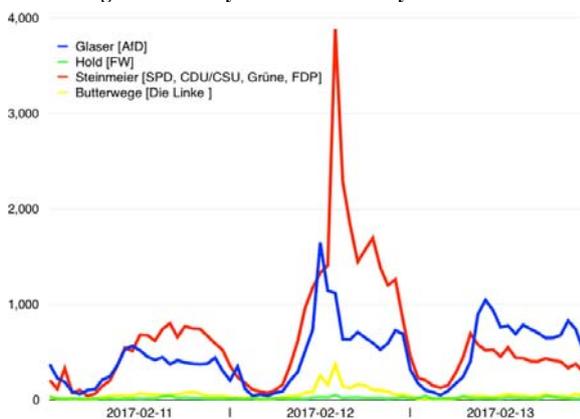
This sample allows us to draw some conclusions about the character and process of political conversation over Twitter during the election. Specifically, we are able to both parse out the amount of social media content related to candidates and affiliated parties and investigate how much of this content is driven by highly automated accounts. Moreover, we can parse the volume of tweets containing URLs, categorize them semi-automatically, identify them as information or misinformation sources and assess the particular contribution of automation to the traffic on this webpage.

Table 1: Twitter Conversation about German Politics around Voting Day, 2017

	N	%
Albrecht Glaser (AfD)	33,125	40.2
Alexander Hold (FW, BVB)	1,064	1.3
Frank-Walter Steinmeier (SPD, CDU / CSU, FDP, Bündnis90 / Die Grünen, SSW)	44,533	54.1
Christoph Butterwegge (Die Linke)	3,627	4.4
Total	82,349	100.0

Source: Authors' calculations from data sampled 11-13 / 02 / 17.
 Note: Glaser hashtags include #afd, #afdwaehlen, #afdwaehlen, #blauswunder, #albrechtglaser, #glaser; Hold hashtags include #hold, #alexanderhold, #fw, #freiewaehler; Steinmeier hashtags include #frankwaltersteinmeier, #steinmeier, #spd, #grüne, #fdp, #cdu, #csu; Butterwegge hashtags include #butterwegge, #christophbutterwegge, #dielinke, #linke.

Figure 1: Hourly Twitter Traffic by Candidate



Source: Authors' calculations from data sampled 11-13 / 02 / 17.
 Note: This figure is based on the hashtags used in the tweets

Table 2: High Frequency Tweeting about German Politics around Voting Day, 2017

	N of Tweets	% of Total	N of Accounts
Albrecht Glaser (AfD)	2,353	7.1	20
Alexander Hold (FW, BVB)	160	15.0	6
Frank-Walter Steinmeier (SPD, CDU / CSU, FDP, Bündnis90 / Die Grünen, SSW)	1,861	4.2	20
Christoph Butterwegge (Die Linke)	279	7.7	16

Source: Authors' calculations from data sampled 11-13 / 02 / 17.

First, we compared the Twitter traffic on the candidates and supporting parties. During the data collection period our script captured 121,582 tweets from 36,541 users. Table 1 reveals the distribution between the hashtags associated with the four major candidates and the parties supporting. The table reveals that the overall volume of traffic on Steinmeier and the parties supporting him (54.1 percent), were dominant on Twitter during the time period investigated. The traffic related to opposition candidate Glaser and supporting party AfD accounted for 40.2 percent. Together Steinmeier and

Glaser traffic accounted for 94.3 percent of traffic. Traffic associated with Hold (1.3 percent) and Butterwegge (4.4 percent) was comparatively small. The overall volume of election-related traffic including one or more general hashtags referring to the election, such as #Bundespräsidentenwahl or #bbpw17, was more than 51,248. Hashtags associated with a general call for political change like #nichtmeinpraesident (not my president) were only used in a small fraction of the election-related tweets (less than 1,000 tweets during the 3 days). Traffic in relation to satirical candidate Engelbert Sonneborn was excluded for the same reason.

Figure 1 displays the rhythm of this traffic over the sample period. There is a significant peak in traffic on Sunday, the day of the election. This is mostly visible in the traffic associated with Steinmeier and the parties supporting him (peaking at 3,881 tweets per hour) and in the Glaser and AfD-related content (peaking at 1,645 tweets per hour). In general, the right-wing opposition party is highly salient in the German Twitter sphere.

Second, we investigated the levels of automation for each candidate. The share of traffic generated by high frequency accounts focusing on the Federal Presidential Elections was not substantial. We identified 22 such accounts, and these accounts generated a total of 5,962 tweets during the 3 days of data collection. This suggests an overall low level of bot-driven automation. These accounts are often bots that see occasional human curation, or they are actively maintained by people who employ scheduling algorithms and other applications for automating social media communication. We define a high level of automation as accounts that post at least 50 times a day, meaning 150 or more tweets on at least one of these hashtags during the data collection period. Our bot detection methodology fails to capture bots tweeting with lower frequencies.

Table 2 reveals that the traffic generated by high frequency accounts with Glaser, Butterwegge and Steinmeier hashtags averaged between 4 and 8 percent. For the Hold-related hashtags, 15 percent of the traffic were automated this way. It is important to note, however, that 150 from the 160 tweets with Hold related hashtags were generated by a single account.

Third, we categorized the sources of information being shared. Table 3 catalogues the different kinds of URLs being shared among people using Twitter to circulate political news and information. Of the tweets sharing URLs captured in this sample, some 14,852 tweets included links to political news and information. To evaluate the qualities and quantities of the various sources of political news and information, we developed a grounded typology with multiple components.

- Professional News Content.

- 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 Political News and Information
 - 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.
- Inaccessible Content.
 - Language. Links that led to content in foreign language that was neither English nor German, when their affiliation could not be verified through reliable sources.
 - 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,

Table 3: What Political News and Information Was Germany Sharing Over Twitter?

Type of Source	N	%	N	%
Professional News Content				
Major News Brands	5,987	89.8		
Minor News Brands	680	10.2		
Subtotal	6,667	100.0	6,667	44.9
Professional Political Content				
Political Party or Candidate	1,543	76		
Government	260	12.8		
Experts	226	11.1		
Subtotal	2,029	100.0	2,029	13.7
Other Political News and Information				
Junk News	1,504	44.8		
Other Political	770	22.9		
Citizen or Civil Society	529	15.7		
Russia	395	11.8		
Humor or Entertainment	113	3.4		
Religion	49	1.5		
Subtotal	3,360	100.0	3,360	22.6
Other				
Social Media Platform	1,978	90.2		
Other Non-Political	215	9.8		
Subtotal	2,193	100.0	2,193	14.8
Inaccessible				
Language	429	71.1		
No Longer Available	174	28.9		
Subtotal	603	100.0	603	4.1
Total	14,852		14,852	100.0

Source: Authors' calculations from data sampled 11-13 / 02 / 17.

or the text used in a UR could be attributed to another source, it is.

Table 3 presents the findings of this grounded catalogue of content. Overall, 44.9% of the political news and information being shared by Twitter users in Germany came from professional news organizations. Links to content produced by government agencies, political parties and candidates, or experts altogether added up to just 13.7% of the total.

The category of “Other Political News and Information” includes many different kinds of content. Two things should be noted across categories. First, the ratio of professional to junk news is roughly four to one. Second, when the amount of junk news is added to Russian-origin news stories, it appears that fully 12.8 percent of all the content that is clearly news and information about politics and the election is of an untrustworthy provenance.

The right-wing, anti-Islam blog *Philosophia Perennis* (179 shares) and the extremist right-wing *Zuerst!* (179 shares) generated the most shares, followed by the similar *Junge Freiheit* (136) and the anti-establishment *Politically Incorrect News* (126).

CONCLUSIONS

The internet has long been a means of manipulating public opinion.¹⁶ Across the discourse in relation to the German Federal Elections and affiliated parties, we find that the proportion of automated accounts tweeting at a frequency of 50 tweets per day or higher is relatively insignificant.

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. We find that in Germany, conversation about politics mirrored the election result in that President elect Steinmeier (54.1%) was the most talked about during the election. The external, right-wing opposition party AfD and their candidate Glaser (40.2%) were almost equally dominant on Twitter. The proportion of traffic that was generated by high frequency bot accounts was not substantial. Social media users in Germany shared many links to political news and information, but professional news outnumbered junk news by a ratio of four to one.

ABOUT THE PROJECT

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

ACKNOWLEDGMENTS AND DISCLOSURES

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

REFERENCES

1. Howard, P. N. & Kollanyi, B. Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum. *ArXiv160606356 Phys.* (2016).
2. Forelle, M., Howard, P., Monroy-Hernández, A. & Savage, S. Political Bots and the Manipulation

- of Public Opinion in Venezuela. *ArXiv150707109 Phys.* (2015).
3. Read, M. Donald Trump Won Because of Facebook. *New York Magazine* (2016).
4. Dewey, C. Facebook Fake-News Writer: ‘I Think Donald Trump is in the White House Because of Me’. *The Washington Post* (2016).
5. Parkinson, H. J. Click and Elect: How Fake News Helped Donald Trump Win a Real Election. *The Guardian* (2016).
6. World Economic Forum. 10. The Rapid Spread of Misinformation Online. *Outlook on the Global Agenda 2014* (2014).
7. Anderson, M., Caumont, A. & comments, rea C. How social media is reshaping news. *Pew Research Center* (2014).
8. Alejandro, J. Journalism in the Age of Social Media. *Reuters Institute for the Study of Journalism* (2014). Available at: <http://reutersinstitute.politics.ox.ac.uk/publication/journalism-age-social-media>. (Accessed: 20th March 2017)
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. Nicola, S. Web Trolls Are Bombarding Angela Merkel With Abuse After Berlin Attack. *Bloomberg.com* (2016).
11. GmbH, F. A. Z. Digitaler Wahlkampf: Frauke Petry und die Bots. *FAZ.NET* (2017). Available at: <http://www.faz.net/aktuell/politik/digitaler-wahlkampf-frauke-petry-und-die-bots-14863763.html>. (Accessed: 26th March 2017)
12. Germany, S. O., Hamburg. Internet-Kommentare von Automaten: AfD will im Wahlkampf Meinungsroboter einsetzen - SPIEGEL ONLINE - Netzwelt. *SPIEGEL ONLINE* Available at: <http://www.spiegel.de/netzwelt/netzpolitik/afd-will-im-wahlkampf-social-bots-einsetzen-a-1117707.html>. (Accessed: 26th March 2017)
13. Bundestagsdebatte: Merkel schimpft über Internet-Trolle. *sueddeutsche.de* (2016).
14. Lypp, D. L. Wirkung von „Social Bots“ ist unter Sach-verständigen strittig. *Deutscher Bundestag* Available at: <https://www.bundestag.de/dokumente/textarchiv/2017/kw04-pa-bildung-forschung-social-bots/488818>. (Accessed: 26th March 2017)
15. 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).
16. Howard, P. N. *Pax Technica: How the Internet of Things May Set Us Free or Lock Us Up*. (Yale, 2015).