

Junk News and Bots during the German Parliamentary Election: What are German Voters Sharing over Twitter?

[COMPROP](#) DATA MEMO 2017.7 / 19 September 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

Automation and propaganda can significantly impact public life 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 German Parliamentary Election for a ten-day period in September 2017. We find that (1) traffic about the far-right Alternative für Deutschland (AfD) accounts for a surprisingly large portion of Twitter activity given that party's share of voter support. (2) The impact of political bots was minor overall, with highly automated accounts generating a small fraction of the Twitter traffic about the election, and most of the bots working in the service of the far-right AfD. (3) Finally, we find that German social media users shared four links to professional news sources for every one link to junk news. Comparing across countries and over time, we demonstrate that this level of professional news consumption is consistently higher than is the case in the US and UK, but lower than in France, and that the level of automation in German Twitter increased only slightly between the Presidential election campaign of February 2017 and the Parliamentary election campaign of September 2017.

SOCIAL MEDIA AND AUTOMATION

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—whether real or 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 a pernicious means of spreading junk news over social networks of family and friends.

Computational propaganda is the automated dissemination of fake news, misinformation, propaganda and other forms of junk news. It 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 are being used to support campaigns of political misinformation on a global scale. During the 2017 German Federal Presidency Election it was found that junk news made up a significant proportion of information shared by users over Twitter. The ratio of links to professional news to junk news was 4 to 1, with right-wing sources being the most shared junk news sources.⁶ The family of hashtags associated with now President Frank-Walter

Steinmeier dominated. Traffic about the far-right Alternative für Deutschland and their candidate Albrecht Glaser accounted for disproportionate Twitter activity given their share of voter support.

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.⁷ Prior research has found that online media promotes less rigorous journalistic practices and favors sensationalist content, regardless of whether the content has been fact checked or is from a reliable source.^{8,9} 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. Both state and non-state political actors 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 German 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?

COMPUTATIONAL PROPAGANDA IN GERMANY

There have been some dramatic examples of computational propaganda in Germany in the weeks before elections. According to *Der Spiegel* the right-wing activist group *Reconquista Germanica* has declared a “war of memes” on the government, using disinformation and bots with the aim of supporting the AfD in the upcoming elections on September 24th of 2017.¹¹ Social bots have repeatedly amplified hashtags that target Angela Merkel and the political establishment, and those that support the AfD.¹² The German satire party *Die PARTEI* has discovered that a network of automated Facebook accounts has steered at least 31 pro-AfD secret Facebook groups.¹³

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 the potential of bots 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.¹⁴

Computational propaganda has emerged as a political issue in Germany, sparking public concerns of voter manipulation and election meddling.¹⁵ In response to the perceived threat, German lawmakers have passed a stringent law, the *Netzwerkdurchsetzungsgesetz*, that imposes fines of up to 50m Euro on social networking companies if they fail to take down defamatory and junk news content. Some experts have warned that such regulations might be overbearing and that they might negatively affect freedom of expression.¹⁶

SAMPLING AND METHOD

Our dataset contains approximately 984,713 tweets generated by 149,573 unique users that were collected between the 1st and the 10th of September 2017, using hashtags associated with the primary political parties in Germany, the major candidates, and the election itself.

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.¹⁷ In order to gather the most complete and relevant data set, we consulted with country experts and used our pilot study data to identify relevant hashtags. 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 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, where 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 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 984,713 tweets, we selected all of the tweets that contained URLs. Between 1–10 September, Twitter users in Germany shared 115,563 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

analyzed. This approach yielded in 11,646 URLs that were then analyzed. Based on a dictionary of classified sources of news and political information from our previous memo on the German Federal Presidency elections we were able to automatically classify 88.9 percent of URLs. A random sample of 10 percent of the rest of the tweets containing URLs was drawn and analyzed.

The classification of each URL was carried out by a team of three coders fluent in the German 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. The grounded typology of news platforms and content types that was used is as follows:

- Professional News Content. This is political news and information by 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 history of professional journalism.
 - Major News Brands. 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
 - Political Party or Candidate. These links are to official content produced by a political party or candidate campaign.
 - Government. These links are to government websites.
 - Experts. This content takes the form of white papers, policy papers, or scholarship from researchers based at universities, think tanks or other research organizations.
- 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.
 - Other Political Content. Myriad other kinds of political content, including sites for buying political paraphernalia, portals like AOL and

Yahoo! that do not themselves have editorial policies or news content, and other forms of political content.

- 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. Includes platforms for civic expression like Change.org and Medium.
 - Russia. This content was produced by known Russian sources of political news and information, such Russia Today and Sputnik.
 - Humor or Entertainment. Links to songs, comedy skits, cartoons and political jokes.
 - Religion. This content was produced by theology organizations and official churches.
- 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 can 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 a foreign language that was neither English nor German, when their affiliation could not be verified through reliable sources.
 - No longer available. Links that were not available at the time of publication.

FINDINGS AND ANALYSIS

First, we compared the Twitter traffic on the candidates and supporting parties. Table 1 describes the ratio between the hashtags associated with the six major candidates and the parties supporting them. The table reveals that Twitter conversations about the AfD and its candidates (30.1 percent of all traffic coded) were dominant on Twitter during the time period investigated. Indeed, the AfD-related volume of tweets was even greater than the neutral election related traffic, which accounted for 29 percent of tweets in the sample.

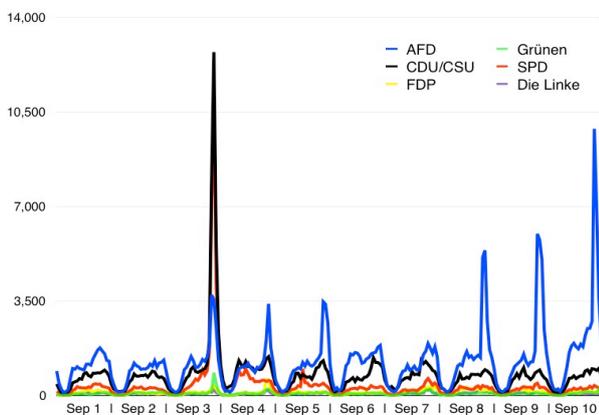
The traffic relating to the CDU/CSU and chancellor Angela Merkel accounted for 18.2 percent of all political traffic. Comparatively, traffic related to the SPD and its candidate Martin Schulz accounted for only 8.9 percent of total Twitter traffic. Traffic associated with the FDP (2.6 percent), Bündnis90/Die Grünen (1.6 percent) and Die Linke (1.5 percent) was comparably unsubstantial.

Figure 1 shows the rhythm of this traffic over the sample period. There is a significant peak in traffic on the 3rd of September, the day of the TV debate

	N	%
General election related	285,185	29.0
AfD	296,658	30.1
CDU/CSU	180,046	18.2
FDP	25,478	2.6
Bündnis90/Die Grünen	15,705	1.6
Die Linke	14,751	1.5
SPD	87,642	8.9
Total	905,465	100.0

Source: Authors' calculations from data sampled 01-10/09/17.
 Note: General election related hashtags include: #btw2017, #bundestagswahl, #wahlkampf. AfD hashtags include: #afd, #holdirdeinlandzurück, #gauland. CDU/CSU hashtags include: #angelamerkel, #fedidwgugl, #CDU. FDP hashtags include: #lindner, #denkenwirneue, #fdp. Bündnis90/Die Grünen hashtags include: #grüne, #darumgrün, #diegrünen. Die Linke hashtags include #dielinke, #linke. SPD hashtags include: #martinschulz, #SPD, #zeitfürmartin. The groups have some overlap as a small number of tweets contained multiple hashtags.

Figure 1: Hourly Twitter Traffic by Candidate



Source: Authors' calculations from data sampled 01-10/09/17.
 Note: This figure is based on the hashtags used in the tweets.

	N of Tweets	N of automated Tweets	% of Total within Party
General election related	285,185	26,821	9.4
AfD	296,658	44,533	15.0
CDU/CSU	180,046	13,099	7.3
FDP	25,478	2,127	8.3
Bündnis90/Die Grünen	15,705	1,752	11.2
Die Linke	14,751	1,819	12.3
SPD	87,642	6,669	7.6
Total	905,465		

Source: Authors' calculations from data sampled 01-10/09/17

between Angela Merkel and Martin Schulz. This is mostly visible in the peak in traffic associated

with CDU/CSU and their candidate Angela Merkel (12,709 tweets per hour) and the peak in traffic associated with the SPD and their candidate Martin Schulz (10,878 tweets per hour). Over the last three days of the period of enquiry, traffic on the right-wing AfD increased substantially. In general, the right-wing opposition party is highly salient in the German Twittersphere.

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 92 such accounts. These accounts generated a total of unique 73,012 tweets during the 10 days of data collection. That is about 7.4 percent of the total traffic. This suggests an overall moderate level of bot-driven automation. We define a high level of automation as accounts that post at least 50 times a day, meaning 500 or more tweets on at least one of these hashtags during the data collection period.

Table 2 reveals that the traffic generated by high frequency accounts for the general election related hashtags, the CDU/CSU, FDP and SPD averaged between 7.3 and 9.4 percent. For the AfD-related hashtags, 15 percent of the traffic was automated this way. Bündnis90/Die Grünen and Die Linke saw 11.2 percent and 12.3 percent of automated traffic, respectively.

Third, we categorized the sources of information being shared. Table 3 catalogues the different kinds of URLs being shared among Twitter users to circulate political news and information. Of the tweets sharing URLs captured in this sample, some 11,646 tweets included links to political news and information. Table 3 presents the findings of this grounded catalogue of content. Overall, 40.2 percent of the political news and information being shared by Twitter users discussing the German election 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 7.4 percent of the total.

The category of “Other Political News and Information” includes many different kinds of content. The ratio of links to professional to junk news is roughly four to one. The junk news sources identified can be distinguished from opinionated content in that they present incorrect information as facts as opposed to opinion. The right-wing, anti-Islam blog *Philosophia Perennis* (156 shares), followed by the conservative, right-extremist *Junge Freiheit* (91).

Fourth, we compared our analysis with our previous findings on computational propaganda. Having performed this analysis over five major elections in the past twelve months, we can now compare the consumption of professional news across several countries and over time. Table 4 shows the levels of automation and junk news shared on Twitter across the major elections that have occurred so far around the world in 2016–2017.

Table 3: German Political News and Information on Twitter

Type of Source	N	%	N	%
Professional News Content				
Major News Brands	4,565	97.6		
Minor News Brands	114	2.4		
Subtotal	4,679	100.0	4,679	40.2
Professional Political Content				
Political Party or Candidate	1,047	85.4		
Experts	99	8.1		
Government	80	6.5		
Subtotal	1,226	100.0	1,226	10.5
Other Political News and Information				
Junk News	1,055	32.2		
Other Political	940	28.7		
Citizen or Civil Society	719	21.9		
Humor or Entertainment	378	11.5		
Russia	130	4.0		
Religion	55	1.7		
Subtotal	3,277	100.0	3,277	28.1
Relevant Content Subtotal			9,182	78.8
Other				
Social Media Platform	1,352	66.2		
Other Non-Political	691	33.8		
Subtotal	2,043	100.0	2,043	17.8
Inaccessible				
No Longer Available	294	69.8		
Language	127	30.2		
Subtotal	421	100.0	421	3.6
Total			11,646	100.0

Source: Authors' calculations from data sampled 01-10/09/17.

Note: General election related hashtags include: #btw2017, #bundestagswahl, #wahlkampf. AfD hashtags include: #afd, #holdirdeinlandzurück, #gauland. CDU/CSU hashtags include: #angelamerkel, #fedidwgugl, #CDU. FDP hashtags include: #lindner, #denkenwirneu, #fdp. Bündnis90/Die Grünen hashtags include: #grüne, #darumgrün, #diegrünen. Die Linke hashtags include #dielinke, #linke. SPD hashtags include: #martinschulz, #SPD, #zeitfürmartin.

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 junk content.

German social media users shared a lower percentage of junk news content than social media users who actively discussed the 2016 US Presidential election in Michigan. They shared a roughly comparable amount of junk news as did UK voters during the 2017 UK General Election, but a higher percentage than voters in the 2017 French Presidential Elections. The ratio of links to professionally produced news to junk news was 4.4:1 for German voters in the run up to the 2017 Parliamentary elections, and has remained constant since the German Federal Presidential Election in February 2017 (see Table 4).

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 France this ranged between 49.4% and 57% of

relevant links across both elections, and in the UK it ranged between 53.4% and 53.6%. In Germany the share dropped from 55.3% during the Federal Presidency Elections in February 2017, to 51.0% during the Parliamentary Election in September 2017.

The 2017 German elections have produced one of the lowest levels of automated content on Twitter as compared to all other democracies we have studied. The level of automation slightly increased from the Federal Election in February to the Parliamentary Election in September with 5.7% and 7.4% of automated content, respectively.

CONCLUSIONS

The internet has long been used to manipulate public opinion.¹² 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 German politics in 2017. We find that in Germany, conversation about politics on Twitter does not mirror the current polls. The right-wing opposition party AfD is dominant on Twitter, with most of the bots in our sample working in their favor. Social media users in Germany have shared many links to political news and information, but links to professional news have outnumbered those to junk news by a ratio of four to one.

ABOUT THE PROJECT

The Project on Computational Propaganda 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. The project gratefully thanks the Ford Foundation for their support. 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 or the University of Oxford.

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	1:1
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 I. Sample, soon after election announced, 1-7 May 2017, 1,363K tweets.	53.4	11.1	12.3	12.6	1.0	1.5:1	4.2:1
United Kingdom II. Sample, candidate debates, 27 May-2 June, 2017, 2.5M tweets.	53.6	10.5	16.5	11.4	1.0	1.5:1	4.7:1
Germany I. Sample, before voting. Federal President Election. 11-13 February 2017, 121K tweets.	55.3	16.8	5.7	12.5	3.3	2:1	4.4:1
Germany II. Sample, before voting. Parliamentary Election. 1-10 September 2017, 985K tweets.	51.0	13.4	7.4	11.5	1.4	1.4:1	4.4:1

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

REFERENCES

- Howard, P., 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).
- Forelle, M., Howard, P., Monroy-Hernández, A. & Savage, S. Political Bots and the Manipulation of Public Opinion in Venezuela. *ArXiv150707109 Phys.* (2015).
- Parkinson, H. J. Click and elect: how fake news helped Donald Trump win a real election. *The Guardian* (2016).
- Read, M. Donald Trump Won Because of Facebook. *New York Magazine* (2016).
- Dewey, C. Facebook Fake-News Writer: 'I Think Donald Trump is in the White House Because of Me'. *The Washington Post* (2016).
- Neudert, L.-M., Howard, P. & Kollanyi, B. Junk News and Bots during the German Federal Presidency Election: What Were German Voters Sharing Over Twitter? (2017).
- World Economic Forum. 10. The Rapid Spread of Misinformation Online. *Outlook on the Global Agenda 2014* (2014).
- Vicario, M. D. *et al.* The Spreading of Misinformation Online. *Proc. Natl. Acad. Sci.* **113**, 554–559 (2016).
- 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).
- Howard, P. N. Digitizing the social contract: Producing American political culture in the age of new media. *Commun. Rev.* **6**, 213–245 (2003).
- Höfner, R., Rosenbach, M. & Hammerstein, M. von. Wahlkampf mit Fake-News: Wie die Rechtsextremen mit Hetze mobil machen. *Der Spiegel* (2017).
- @DFRLab. #BotSpot: Memes Target Der Spiegel, Merkel. *DFRLab* (2017).
- Locker, T. Was hinter den Bots steckt, die die 31 geheimen AfD Gruppen gesteuert haben. *Motherboard* (2017). Available at: <https://motherboard.vice.com/de/article/qvv7eb/was-hinter-den-bots-steckt-die-die-31-geheimen-afd-gruppen-gesteuert-haben>.
- Stern, J. AfD verzichtet auf Meinungsroboter - oder nicht? *tagesschau.de* (2017). Available at: <http://www.tagesschau.de/multimedia/video/video-283363~player.html>. (Accessed: 15th September 2017)
- Neudert, L. M. *Computational Propaganda in Germany: A Cautionary Tale*. 31 (Oxford Internet Institute, University of Oxford, 2017).
- Krempel, S. Netzwerkdurchsetzungsgesetz: Experten haben gravierende verfassungsrechtliche Bedenken. *heise online* (2017). Available at: <http://www.heise.de/newsticker/meldung/Netzwerkdurchsetzungsgesetz-Experten-haben-gravierende-verfassungsrechtliche-Bedenken-3747671.html>. (Accessed: 15th September 2017)
- 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)