

Data Descriptor

The #BTW17 Twitter Dataset—Recorded Tweets of the Federal Election Campaigns of 2017 for the 19th German Bundestag

Nane Kratzke 

Center of Excellence for Communication, Systems and Applications (CoSA), Lübeck University of Applied Sciences, 23562 Lübeck, Germany; nane.kratzke@fh-luebeck.de

Received: 25 September 2017; Accepted: 18 October 2017; Published: 20 October 2017

Abstract: The German Bundestag elections are the most important elections in Germany. This dataset comprises Twitter interactions related to German politicians of the most important political parties over several months in the (pre-)phase of the German federal election campaigns in 2017. The Twitter accounts of more than 360 politicians were followed for four months. The collected data comprise a sample of approximately 10 GB of Twitter raw data, and they cover more than 120,000 active Twitter users and more than 1,200,000 recorded tweets. Even without sophisticated data analysis techniques, it was possible to deduce a likely political party proximity for more than half of these accounts simply by looking at the re-tweet behavior. This might be of interest for innovative data-driven party campaign strategists in the future. Furthermore, it is observable, that, in Germany, supporters and politicians of populist parties make use of Twitter much more intensively and aggressively than supporters of other parties. Furthermore, established left-wing parties seem to be more active on Twitter than established conservative parties. The dataset can be used to study how political parties, their followers and supporters make use of social media channels in political election campaigns and what kind of content is shared.

Data Set: <https://doi.org/10.5281/zenodo.835735>

Data Set License: CC BY 4.0

Keywords: Twitter; dataset; Bundestag; election campaign; Germany; 19th German Bundestag

1. Introduction

Data-driven political campaigns can be successful. “The Obama 2012 campaign used data analytics and the experimental method to assemble a winning coalition vote by vote. In doing so, it overturned the long dominance of TV advertising in U.S. politics and created something new in the world: a national campaign run like a local ward election, where the interests of individual voters were known and addressed” [1]. However, four years later, Hillary Clinton’s data-driven campaign organized by the same party failed under the eyes of the world [2]. The question is why data-driven campaigns worked for Barack Obama, but not for Hillary Clinton?

Both campaigns focused on targeting a specific and very small group of citizens based on sociodemographic and psychographic data. However, it seems also important to understand the network that connects to the candidates and how many nodes of this network can be reached and engaged in political interactions to distribute a political message or vision. It is interesting that one target of the Trump campaign was not to mobilize its own supporters, but to demobilize Clinton supporters in order to deactivate part of the political opponent’s network. It should be obvious for the reader that data that have been collected during such election campaigns might contain

valuable insights worth being mined by the political science and the political communication research community. Interesting questions arise about how the analysis of social media channels can be used more systematically by political scientists and and political communication analysts.

- What are the main influencers and multipliers in this political network?
- Are these influencers and multipliers aware or unaware in established political communication research?
- Are the influencers and multipliers influenceable?
- How robust are influencer and multiplier networks against disturbance of trolls and demobilizing effects?
- Is it possible to identify groups of multipliers that can be influenced more easily than other groups?
- Is it possible to identify relevant societal trends from social media streams that might be not covered by the political communication sufficiently?
- Is it possible to make use of social media as an early-warning system for rising societal trends that political actors are currently unaware of?
- Is it possible to identify commonalities of groups that feel politically penalized and misunderstood?
- Is it possible to measure this feeling of being politically penalized and misunderstood?
- And so on.

Twitter data analysis gets more and more common for these kinds of questions in the (social) sciences and is, beside other domains, applied to understand the influence of social media on democratic election campaigns. Barberá and Rivero emphasize “the opportunities offered by Twitter for the analysis of public opinion: messages are exchanged by numerous users in a public forum and they may contain valuable information about individual preferences and reactions to different political events in an environment that is fully accessible to the researcher” [3].

Furthermore, Twitter provides samples of these data for free via its streaming APIs. At least for large datasets [4], these samples “truthfully reflect the daily and hourly activity patterns of the Twitter users (...) and preserve the relative importance (...) of content terms” [5]. Twitter might not be the biggest, but one of the most influencing social networks providing a microblogging service with more than 320 million active Twitter users in 2017. Compared with other data, the provided dataset is small, and its content is more focused (compared with other sources like Facebook). Twitter data have been used for a variety of interesting studies.

- Analysis of the political representativeness of Twitter users (Spanish election campaign of 2011 and U.S. presidential election campaign of 2012) [3]
- Twitter status updates in the context of Live-TV events [6]
- Tweets and Votes, a Special Relationship: The 2009 Federal Election in Germany [7]
- Real-time Twitter sentiment analysis of 2012 U.S. Presidential election cycle [8]
- Limits of Electoral Predictions using Twitter [9]
- Social media adaption in the U.S. congress [10]
- Twitter adoption and activity in U.S. legislatures [11]
- The uses of Twitter by populist presidents in contemporary Latin America [12]

Furthermore, there exist several Twitter datasets with a clear focus on political election campaigns in countries of the European Union [3,7,13,14]. This dataset has been collected to provide data for one further European country (Germany). The methods for how the data have been collected are summarized in this paper. Compared with the previously-mentioned studies and datasets (except [13]), this dataset is bigger and comprises more than 10 GB of raw data, more than 120,000 observed Twitter accounts and more than 1,200,000 recorded tweets. However and in contrast with [13], the collection method did not strive for an intentionally large dataset and did not try to cover the complete German Twitter data stream.

One major motivation to record this dataset was to collect Twitter data in the “hot” (pre-)phase of political election campaigns in Germany. So far, it is obvious that German parties make use of social media, but not on a comparable level like the U.S. campaigns in 2008, 2012 and 2016 or the U.K. Brexit campaigns in 2016. However, it is more than likely that the professionalism (or the “data-drivenness”) will increase in the future. Therefore, this dataset might be one of the latest datasets without being affected by too many social media effects in political campaigning in Germany [15].

Therefore, this dataset might be used as a reference dataset for future studies that want to study the long-term effects of social media on political campaigning. Furthermore, it can be used to test or investigate hypotheses regarding several social media-related phenomena of modern election campaigns like:

1. The design of future political campaigns using social media more systematically.
2. Mechanisms of hate-speech, populism and their correlated network structures.
3. Classification of Twitter accounts regarding political party proximity.
4. Identification of influencing and multiplying Twitter accounts.
5. Identification of strength and weaknesses in network structures that shall be considered for effective political communication through social media channels.
6. Understanding motivations to distribute political content (more than 50% of all Twitter interactions are re-tweets).
7. The effectiveness to identify swing-voters (which are only 3% according to this dataset).
8. The limitations to target specific citizens (most observed Twitter users are too inactive to do this).
9. And so on.

Last, but not least, the dataset provides more than 1,200,000 tweets in German that can be used to develop, test and improve natural language processing and machine learning tools for the German language. More details on possible applications, but also on limitations and ethical considerations of this dataset can be found in Sections 4.1, 4.2, 4.3 and 4.4.

2. Data Acquisition and Processing Including Quality Control Measures

The data acquisition has been done according to the approach described by [14] using the Twitter streaming API filtering all tweets concerning a list of Twitter accounts belonging to relevant German politicians. The recording started on 29 May 2017 and ended on 24 September 2017 (the election day). Relevant Twitter accounts of politicians have been selected in a two-step and semi-automatic process.

1. In a first step, the official websites of the 18th German Bundestag factions have been crawled for Twitter screen names, because these websites contain lists of politicians with links to their official social media accounts.
2. In a second step, these resulting screen names were checked for plausibility to exclude wrong screen names. Some pages contain Twitter live tweets of politicians containing screen names out of the political context like @Sportschau (famous German TV show for sport). This step has been done manually. In rare cases, some Twitter accounts were added like @MartinSchulz (one of the chancellor candidates for the 19th German Bundestag). Due to his former membership of the European Parliament¹, he was not a member of the 18th German Bundestag, but was obviously a relevant actor in the political discussion.

Because the Alternative für Deutschland (AfD) and the Freie Demokratische Partei Deutschlands (FDP) were no members of the 18th German Bundestag (but it was likely that they will enter the 19th

¹ Martin Schulz was the President of the European Parliament before being nominated as chancellor candidate by the Social Democratic Party (SPD).

German Bundestag), further official websites of these parties were selected to crawl for relevant and representative Twitter accounts of politicians. In the case of the AfD, this was the website of the directorate of the AfD federal party and the list of members of the European Parliament; in the case of the FDP, this was the website of the executive committee of the FDP federal party of Germany. Table 1 shows all crawled websites to identify Twitter accounts of relevant politicians. Appendix A lists and groups the Twitter screen names of all followed politicians. These accounts were checked to be valid and official accounts to avoid possible sources of error and noise.

Table 1. Relevant Twitter accounts of politicians were crawled from official faction websites. The Freie Demokratische Partei Deutschlands (FDP) and Alternative für Deutschland (AfD) sites have been taken additionally into consideration because it was likely that both parties will enter the 19th German Bundestag. CDU, Christian Democratic Union; CSU, Christian Social Union; SPD, Social Democratic Party.

Faction	Website	Followed Depth
CDU/CSU	https://www.cdusu.de/abgeordnete	0
SPD	http://www.spdfraktion.de/abgeordnete/alle	0
Linke	https://www.linksfraktion.de/fraktion/abgeordnete/a-bis-e	1
	https://www.linksfraktion.de/fraktion/abgeordnete/f-bis-j	1
	https://www.linksfraktion.de/fraktion/abgeordnete/k-bis-o	1
	https://www.linksfraktion.de/fraktion/abgeordnete/p-bis-t	1
	https://www.linksfraktion.de/fraktion/abgeordnete/u-bis-z	1
Grüne	https://www.gruene-bundestag.de/abgeordnete.html	0
FDP	https://www.fdp.de/seite/praesidium	0
AfD	https://www.afd.de/partei/bundesvorstand/	0
	https://www.afd.de/partei/eu-abgeordnete/	0

3. Dataset Description

This dataset description provides no content-based (in-depth) analysis of observed Twitter interactions. This will be done by follow-up analysis in the aftermath of data collection. The dataset is just described in a descriptive and quantitative manner to provide the reader some hints about possible research directions, but also limitations. Table 2 lists and describes the considered political parties of this dataset. Three hundred twenty eight Twitter accounts of all 364 observed accounts were active accounts (that means at least one Twitter interaction has been recorded during the period of observation).

The dataset comprises more than 1,200,000 tweets from 120,000 users. These recorded tweets and users are stored in exactly the JSON-based API (raw) format provided by the public Twitter streaming API (see Appendix B). The dataset contains:

- screen names of observed politicians and their party membership,
- texts of recorded Tweets,
- user information of observed users provided by the Twitter Streaming API at the time of recording,
- user mentions,
- hashtags,
- media and further references,
- identifiers of users and tweets to query additional information per Tweet or Twitter user,
- interactions between users (reply to tweet, quote of tweet, re-tweet of tweet),
- and timestamps for all status posts and Twitter interactions (replies, re-tweets, quotes).

Table 2. Observed Twitter accounts in numbers and parties. Party descriptions are taken and adapted from the English version of Wikipedia. Crawled Twitter accounts are accounts with links from official faction websites of the 18th German Bundestag. However, that must not mean that these accounts are used actively to disseminate political content or to engage in political discussions on Twitter. Active accounts are Twitter accounts that sent at least one tweet in the period of data collection. FDP and AfD were considered because it was likely that both parties will enter the 19th German Bundestag although not being part of the 18th German Bundestag.

Party	Twitter Accounts Crawled	Twitter Accounts Active	Seats in 18th Bundestag	Description
CDU/CSU	105 (34%)	93 (89%)	309	The Christian Democratic Union of Germany is a Christian democratic (German: Christlich Demokratische Union Deutschlands) and liberal-conservative political party in Germany. It is the major catch-all party of the center-right in German politics. The CDU forms the CDU/CSU faction, also known as the Union, in the Bundestag with its Bavarian counterpart the Christian Social Union in Bavaria (CSU).
SPD	86 (45%)	84 (98%)	193	The Social Democratic Party of Germany (German: Sozialdemokratische Partei Deutschlands, SPD) is a social-democratic political party in Germany. The party is one of the two major contemporary political parties in Germany, along with the Christian Democratic Union (CDU).
Linke	60 (94%)	49 (82%)	64	The Left (German: Die Linke), also commonly referred to as the Left Party, is a democratic socialist and left-wing populist political party in Germany. The party was founded in 2007 as the merger of the Party of Democratic Socialism (PDS) and the Electoral Alternative for Labour and Social Justice (WASG).
Grüne	62 (98%)	56 (90%)	63	Alliance 90/The Greens, often simply Greens, (German: Bündnis 90/Die Grünen or Grüne) is a green political party in Germany, formed from the merger of the German Green Party (founded in West Germany in 1980) and Alliance 90 (founded during the Revolution of 1989–1990 in East Germany) in 1993. The focus of the party is on ecological, economic and social sustainability.
FDP	39 (-)	36 (92%)	0	The Free Democratic Party (German: Freie Demokratische Partei, FDP) is a classical liberal political party in Germany. In the 2013 federal election, the FDP failed to win any directly-elected seats in the Bundestag and came up short of the 5 percent threshold to qualify for list representation. The FDP was therefore left without representation in the Bundestag for the first time in its history.
AfD	12 (-)	10 (83%)	0	Alternative for Germany (German: Alternative für Deutschland, AfD) is a right-wing populist and Eurosceptic political party in Germany founded in 2012/2013. The AfD was founded as a center-right conservative party of the middle class with a 'soft' Euroscepticism, being generally supportive of Germany's membership in the European Union, but critical of further European integration, the existence of the euro currency and the bailouts by the Eurozone for countries such as Greece. Over the years, the party has become more and more nationalistic.

However, not all of these data are used for this dataset description. The dataset is mainly described by providing the following descriptive quantitative data concerning:

- volume of tweets during the political campaigns for the 19th German Bundestag,

- percentages of tweet subtypes (status messages, re-tweets, replies, quotes),
- amount of engaged Twitter users per party,
- and party-specific observations of account ages and “loudness” of re-tweeting Twitter accounts.

3.1. Volume of Tweets during Political Campaigns

Figure 1 shows the amount of tweets over time for the observation period. The amount of tweets by politicians ranges between 2000 tweets and 10,000 tweets a day. The maximum was reached at the days:

- of the ballot on same-sex marriage in German Bundestag on 30 June 2017,
- around TV debates on 3–5 September 2017 and
- on the election day of 24 September 2017.

The total amount of tweets ranges between 5000 and more than 35,000 tweets a day (see Figure 1).

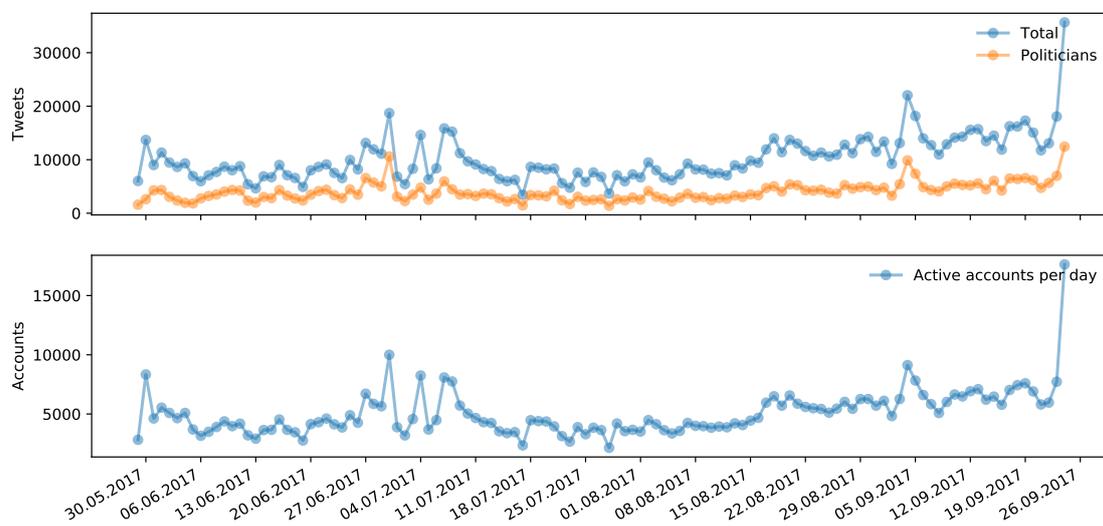


Figure 1. Observed tweets and active accounts over time. Very often, peaks indicate major election campaign events like TV debates around 3–5 September or the election day on 24 September 2017.

3.2. Percentages of Tweet Types

Figure 2 shows the relation of tweet types (public status messages, replies, re-tweets or quotes) and indicates a quite constant relation of tweet types over time. According to observed tweets:

- half of all tweets are re-tweets,
- a third of all tweets are replies,
- a tenth of all tweets are quotes (which is a kind of re-tweet, but adds additional content or context that may change the intended message of the original tweet),
- and only 5% of all tweets are status messages (containing political content or statements).

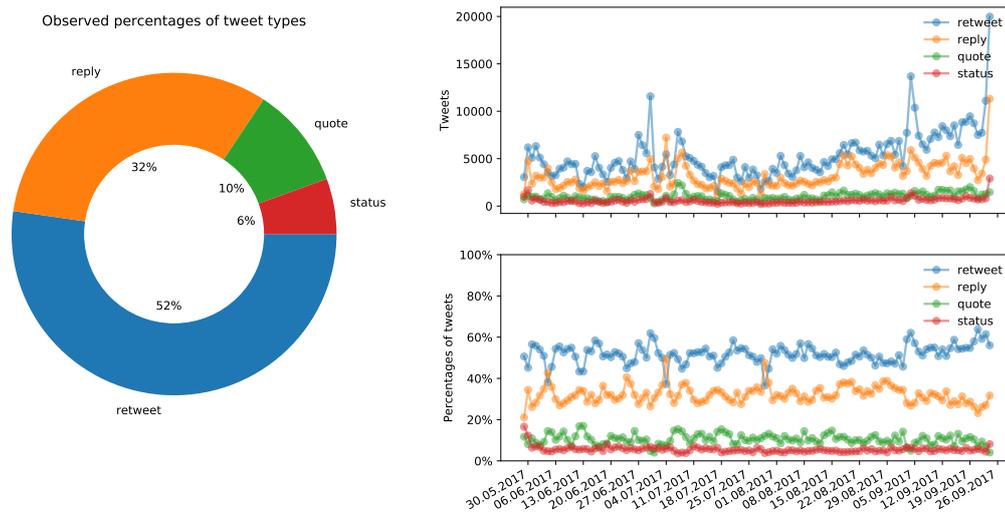


Figure 2. Volume of tweet types. The relation of tweet type percentages stays more or less constant over time although the volume of tweets, re-tweets, quotes and replies increases until election day (24 September 2017).

In other words, 94% of all tweets are a discussion about or further dissemination of political content, and only 6% is political content. Because almost half of all recorded tweets were re-tweets, these re-tweets were used to group observed Twitter users. A re-tweet is a good indicator of whether a user conforms to a political position. If a user re-tweets a politician, it can be assumed that the user conforms with the tweet. However, all other kinds of interactions are not so clear.

- A quote is a kind of a re-tweet, but it adds additional content that might change the intended message of the original tweet. Whether the intended message is questioned or not must be analyzed using the content of the tweet. For that kind of analysis, natural language processing (NLP) can be applied. However, this kind of information is not used to build groups of users for this dataset description.
- A reply might be supporting, contradicting or simply questioning. However, that can be only determined by content-based text classification or other natural language processing and analyzing techniques. Therefore, this kind of information is not used to build descriptive groups of users for this dataset description.
- A status post (maybe mentioning a political actor) might mean everything (supporting, contradicting, questioning, spoofing or just mentioning and much more). Furthermore, for this kind of tweet, it is necessary to evaluate the tweet on a content basis. Therefore, this kind of information is not used either. Furthermore, the reader should remember that only 5% of all observed Tweets were status posts. Therefore, evaluating status posts might not be worth the effort (although this might sound contradictory at first).

It is important to understand that this dataset description groups according to re-tweets because it is objective and considers efficiently more than 50% of all observed Twitter interactions; however not all interactions and data are considered. If an analyst wants to derive a “true” political party proximity, other valuable information sources should be taken into consideration. This could be a follower graph or a sentiment analysis of tweets as done by other studies [16,17]. This dataset description uses re-tweets because it is efficient and sufficient for a first and descriptive dataset description. The author does not proclaim that the single focus on re-tweets is sufficient for an in-depth analysis.

3.3. Re-Tweeting Twitter Users per Party

Therefore, only re-tweets were applied to group observed Twitter users for this dataset description. This was done according to the following scheme.

- If a Twitter user re-tweets mostly tweets of one political party, this user is assigned to that party. This group is likely to contain a higher-than-average amount of voting for the respective party.
- A user is assigned to the group ‘inconsistent’ if the user re-tweets tweets of more than one political party. This group may contain so called swing-voters. However and according to Figure 3, this group is quite small, and an in-depth analysis shows that many these kinds of accounts are newspapers, radio or TV-stations that try not to re-tweet disproportionately high content of a specific party.
- If a Twitter user re-tweets no tweets of any political party, this user is assigned to the group ‘unknown’. This group contain voters where little can be derived from the re-tweeting behavior. Figure 3 shows that approximately 45% of all observed Twitter users can not be assigned to a political party simply by looking at their re-tweeting behavior. This share of “unknowns” might be reducible by applying more sophisticated natural language processing-based analysis of quotes, replies and status messages (which was not done for this dataset description).

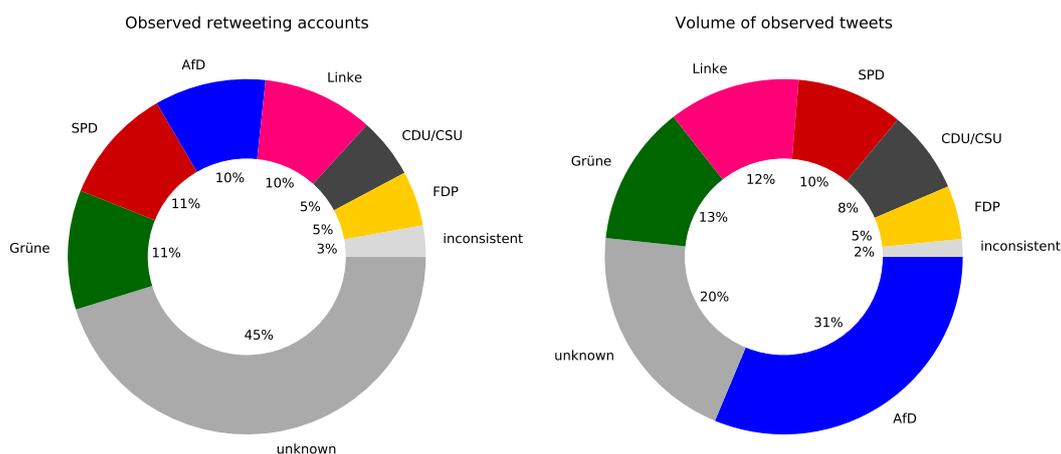


Figure 3. Observed party proportion in accounts and tweets. These pie charts visualize the percentages of party content re-tweeting users (**left**) and the percentages of associated tweet volumes by these users (**right**). If the proportions do not sum up to 100% this is a presentation error due to rounding.

Figure 3 shows the result of this grouping. Left wing parties like SPD, Grüne and Linke seem to have a bit more re-tweeters compared with conservative or liberal parties like CDU/CSU and FDP. The right-wing populist AfD is somewhere in between. However, if we consider the tweet volume, it can be seen that almost for all parties, the amount of tweets relates more or less to the amount of re-tweeting users. Only the right-wing populist AfD has substantially more re-tweets than all other parties (almost 30% of all observed tweets were generated by only 9% of observed Twitter users re-tweeting AfD content). If it is taken into account that far less AfD politicians were observed than from any other party, this becomes even more astonishing (see Table 2).

3.4. Account Ages of Party Re-Tweeters

Therefore, re-tweeters of the right-wing populist AfD seem to be “louder” than other re-tweeters. To be “louder” means the same amount of re-tweeters generate much more Twitter interactions (replies, re-tweets and quotes). That might have several reasons: for instance, the age of the Twitter

accounts. To make a new Twitter account known, this account should publish or distribute content. As a consequence younger, Twitter accounts have the tendency to be more active than older, more established accounts. Figure 4 shows the histogram of account ages² per party. The left side shows the absolute numbers of accounts. The conservative CDU/CSU and liberal FDP are the parties with the fewest re-tweeting accounts. The (middle-)left SPD, Grüne and Linke have more re-tweeting accounts, and the right-wing populist AfD shows a sharp increase of accounts in the last two or three years. The rise of accounts is visualized much more clearly on the right-hand side of Figure 4 where the relative distributions of account ages per party are shown. It is clearly observable that a third of all AfD re-tweeting accounts are younger than one year, and almost 50% of re-tweeting accounts did not exist two years before. A similar, but not so distinctive, effect can be observed for the liberal FDP. Neither party had any seats in the 18th German Bundestag and therefore smaller media presence. Both parties seem to be compensating that by making more intensive use of social media channels like Twitter. This strategy seem to work especially for Twitter users who have not been on Twitter long. Maybe these accounts have been even created intentionally to support the social media efforts. Similar effects could be observed four and especially eight years ago, as well (campaigns for the 18th and 17th German Bundestag). Furthermore, the group of “unknowns” shows that especially younger accounts are engaged in political Twitter interactions.

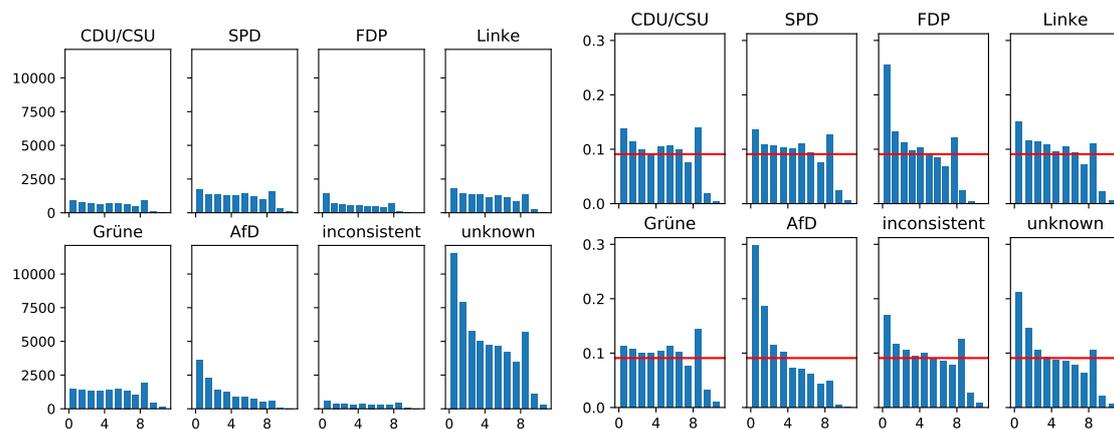


Figure 4. Ages of observed accounts that re-tweet party content in absolute (**left**) and relative numbers (**right**). It is plotted how old Twitter accounts are (how many years ago the Twitter account has been created). The red line indicates the statistical expectation if account creations were equally distributed over the years.

3.5. “Loudness” of Party Re-Tweeters

Re-tweeters of FDP, AfD and unknowns share some similarities in their account ages, but is this aligned with different observable Twitter usage patterns? The reader already observed that a substantial amount of re-tweets was generated by AfD re-tweeters (see Figure 3). Therefore, the AfD seems to be “louder” than other parties on Twitter. Figures 5–8 visualize the loudness of all party re-tweeters. A histogram of all observed status posts, replies, re-tweets and quotes is presented in gray. On the *x*-axis is plotted how often a tweet type was posted, and on the *y*-axis is plotted how many accounts did this. The blue proportion shows exactly the same, but only for the group of party re-tweeters. These figures can be used to visualize the “loudness” of a specific group of re-tweeters. The more blue a histogram is, the “louder” the specific group is.

² It is important to note that it is not shown how old the owner of the Twitter account is. It is shown how old the Twitter accounts since creation are.

- Taking Figure 5, we see that only users with unknown party proximity tend to post slightly more status posts, but the status posting behavior seems quite comparable across all groups.
- Figure 6 shows that the reply behavior of re-tweeters of established parties like CDU/CSU, SPD, FDP, Grüne and Linke is comparable. However, re-tweeters of the right-wing populist AfD and the group of “unknowns” seem to dominate the reply space. It would be interesting to analyze whether these replies correlate with the postulated “hate-speech” phenomenon that was criticized mainly by established political parties during the election campaigns.
- Figure 7 visualizes the quoting loudness and clearly indicates that quoting is substantially done by re-tweeters of the right-wing populist AfD. It would be interesting to determine whether quoting is used to discredit systematically other political positions. This would be a behavior that is meant to be specific for populist parties.
- Figure 8 shows that re-tweets are applied disproportionately high by the right-wing populist AfD and the group of “unknowns”. Left-wing parties like SPD, Linke and Grüne seem to have slightly more re-tweets than conservative or liberal parties like CDU/CSU and FDP. Furthermore, it would be interesting to analyze what kind of tweets are re-tweeted more often. That would be valuable for parties in order to curate attractive political content.

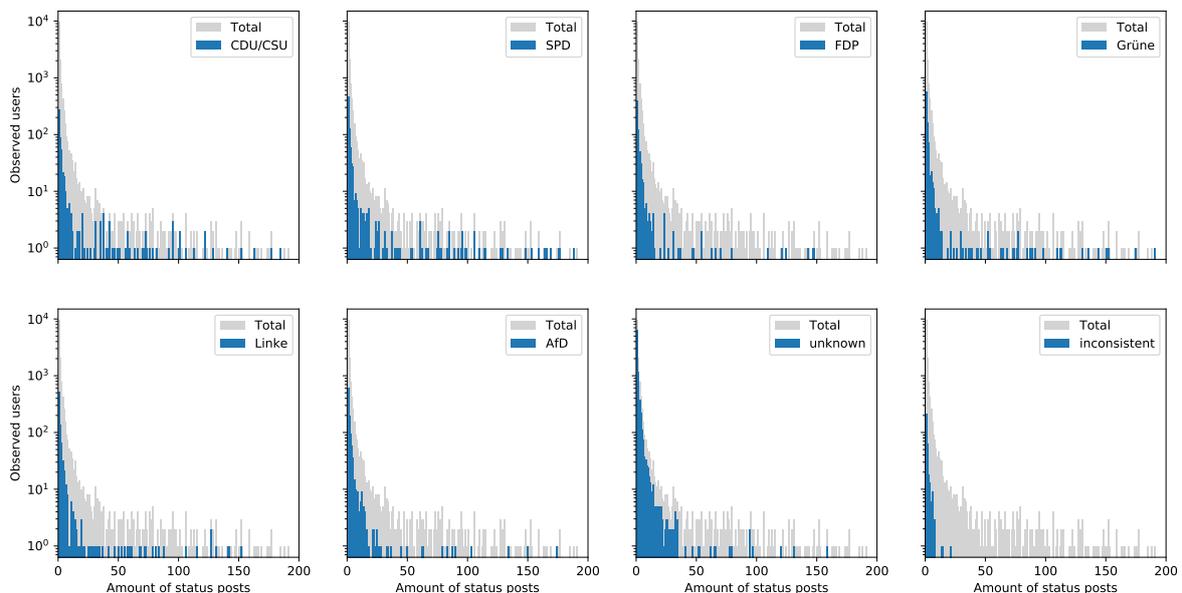


Figure 5. Loudness of status posts by users re-tweeting party content. Users with unknown party proximity tend to post slightly more status posts than other users. However, there seems to be no significant difference between the groups.

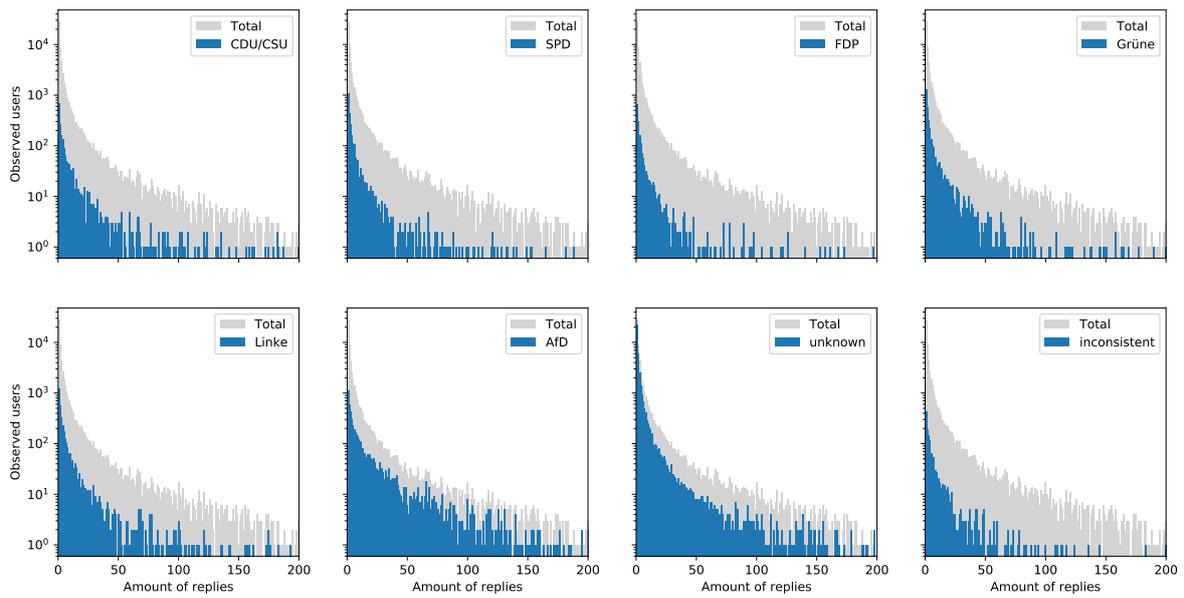


Figure 6. Loudness of replies by users re-tweeting party content. Users with unknown party proximity or AfD proximity (right-wing populist) tend to post more replies than other users.

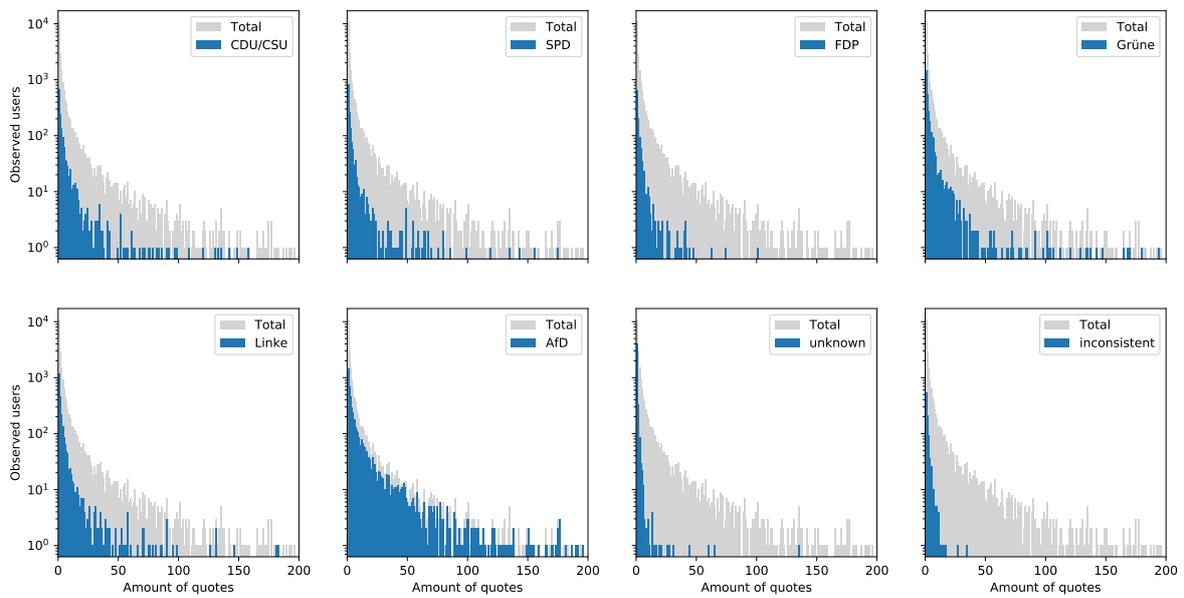


Figure 7. Loudness of quotes by users re-tweeting party content. Users with AfD proximity (right-wing populist) tend to quote more tweets than users of other groups.

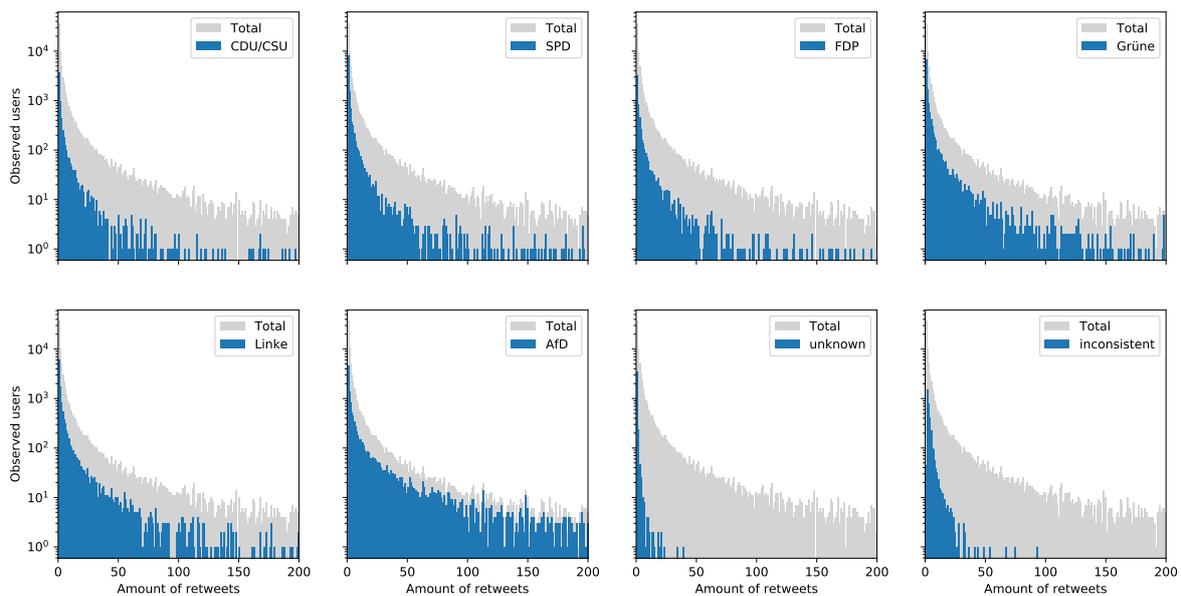


Figure 8. Loudness of re-tweets by users re-tweeting party content. Users with AfD proximity (right-wing populist) tend to re-tweet more tweets than users with a proximity to left-wing parties like Grüne or Linke.

4. Data Use and Application

The dataset might be used by election campaign strategists, political analysts or scientists. Nevertheless, to draw valid (or legal) conclusions, some limitations and ethical considerations should be mentioned.

4.1. Limitations due to Twitter User Protection Terms and Ethical Considerations

The Twitter User Protection terms of use must be respected³. The provided data may not be used for surveillance purposes like tracking, alerting or other monitoring. The data may not be used to conduct surveillance or gather intelligence with the primary purpose to isolate a group of individuals or any single individual for any discriminatory purpose. The data may not be used to target, segment or profile individuals due to their political affiliation or any other category of personal information.

Furthermore, ethical considerations around the re-use of Twitter data should be considered. The dataset contains a sample of Twitter interactions by Twitter users as raw data (this is the same amount of data and the same structure as provided by the Twitter Streaming API). These interactions are intentionally not connected with any computable attribute like party preferences, because this is problematic from an ethical viewpoint. Party preferences are clearly and obviously very personal data. However, the dataset to derive such a kind of preference as this would be possible if an analyst were to use the Twitter Streaming API directly. Therefore, to use this dataset, the same ethical considerations must be made by an analyst as if he/she would make use of the Twitter Streaming API directly.

This dataset intends to analyze re-tweeted parties, but not the re-tweeters. Theoretically, it is possible to derive a political preference of a specific Twitter account using algorithms like label propagation, but this dataset is not really useful for such a kind of analysis; mainly because it is ethically questionable and against the Twitter User Protection terms. However, the following aspects should be considered, as well.

³ See <https://dev.twitter.com/overview/terms/agreement-and-policy> (especially Part VII. Other Important Terms; A–User Protection).

- A Twitter account does not automatically belong to a physical person. It is not unlikely that Twitter accounts belong to a company, an organization or is operated by staff of social media experts on behalf of a person of public interest or organizational function.
- To draw any conclusions for specific accounts is for the vast portion of this dataset statistically questionable. For instance, more than 75% of all observed users re-tweeted less than three times. To derive a party preference for a specific account on so few data is questionable. The dataset intends to enable deriving conclusions for the re-tweeted party, but not for specific re-tweeters. The reader should consider the following Section 4.2 to consider such statistical limitations.

To sum these ethical considerations up, Twitter interactions happen in the open space, and every Twitter user is aware of that by accepting the Twitter terms and conditions. To protect Twitter users' privacy, this dataset does not contain tweets that were not intended for the public like private messages, for instance. Therefore, only the raw data of public tweets are provided to enable a maximum of analytical use cases. However, it was decided against enriching this dataset with further computed attributes like party preferences. Any computed attribute is up to interpretation. Therefore, to compute any kind of computed descriptive attribute should be up to analysts and their particular analytical questions of interest. To do this, the analysts have to consider the Twitter User Protection Terms and relevant ethical standards. This is out of the scope of this dataset and the responsibility of the analyst. Because of these considerations, there is no tool provided with this dataset that could be easily used to derive any (and likely) misleading conclusions for specific Twitter accounts.

4.2. Statistical Limitations to Determine the Party Proximity of Specific Accounts

As the analysis of the “loudness” of party re-tweeters shows, a huge proportion of the data cannot be used to determine the party proximity of a specific account. Figure 8 shows that for most accounts, only very few re-tweets (in many cases, only one or two) were observed. The same is true for replies, quotes and status posts. While this is enough data to reason about how many accounts in sum may have a party proximity, only one or two re-tweets are obviously too few to speculate about the party proximity of a specific account. Therefore, the dataset should not be taken to classify the party proximity of specific observed Twitter accounts. This is especially true for accounts that are on the left-hand side of the histograms being visualized in Figures 5–8.

4.3. Technical Recording Limitations for the Analysis to Be Considered

The Twitter streaming API returns only a sample of all tweets flowing through the Twitter social network. Data analysis must consider this and should take corresponding studies into consideration [4]. It is not assured by Twitter how big this sample size is. However, Twitter states a sample size range between 1% and 10% for tweets. Studies that measured this sample size reported a sample size between 0.95% and 9.6% for tweets and between 10% and 45% for users [4,5]. Wang et. al. concluded that “the sample datasets truthfully reflect the daily and hourly activity patterns of the Twitter users. (...) Even with a very small sampling ratio (i.e., 0.95%), the sample datasets (...) preserve the relative importance (i.e., frequency of appearance) of the content terms” [5].

Furthermore and due to applied filters on the Twitter Streaming API, only re-tweets, tweets, replies or quotes referencing at least one of the accounts listed in Appendix A were recorded systematically. However, that might not be all of the relevant data for specific questions of interest. This effect might be observable especially in political large-scale events like the TV debate of the chancellor candidates (Angela Merkel, CDU, and Martin Schulz, SPD). The Twitter hashtag for this TV event was established as #TVDuell, and many users used this tag, but did not reference @MartinSchulz or Angela Merkel (which had no Twitter account at the time of recording the dataset) in their tweets explicitly. Therefore, these tweets were not recorded and are not part of the sample. Because relevant hashtags could not have been anticipated completely at the start of observation, these kinds of tweets were not recorded systematically. This is likely no problem for the complete

observation time frame, but the dataset might not be adequate to draw conclusions for short-term events of just some hours like the chancellor candidate's TV debate on 3 September 2017. The reader might want to check [4,6] for better problem awareness.

4.4. How This Dataset Can Be Used

The intended purpose of this dataset is to enable the analysis of the 2017 election campaigns for the 19th German Bundestag. The kinds of analysis and questions to be investigated are not limited and are fully up to the data scientist using this dataset. The above-mentioned limitations should be taken into consideration carefully.

However, the key for successful and advanced studies is likely not to focus on the re-tweeting behavior alone. On the one hand, focusing on re-tweeting interactions is simple and can be done without sophisticated supervised or unsupervised machine learning algorithms or natural language processing techniques analyzing the content of tweets. Furthermore, the results of machine learning algorithms can be hard to interpret, and natural language processing or sentiment analysis is often optimized for the English language (but this a German dataset). On the downside, 50% of observed interactions are not used. Therefore, more advanced studies of the dataset should focus the content of tweets, quotes and replies using natural language processing (NLP) frameworks like natural language toolkit (NLTK) [18]. Furthermore, advanced studies should focus on the network structures, as well. These network structures can be built by analyzing the recorded 1,200,000 interactions of the 120,000 observed Twitter users using frameworks like NetworkX [19]. Just to give the reader some guidance, some nearby investigations according to this pre-analysis could be:

- The dataset can be used to understand the mechanisms of hate-speech, populism and the correlated network structures better. It might be even used as training data for social media providers to identify German hate-speech more accurately.
- The dataset can be used to study how Twitter interactions like replies and quotes could be used to optimize the classification of Twitter accounts. This will likely involve the application of machine learning and NLP measures.
- Analyzing the network structure might help to identify influencing and multiplying Twitter accounts. These kind of accounts might be of particular interest for social media campaign strategists.
- The analysis of the strength and weaknesses in network structures can be analyzed to be considered effective in political communication.
- Election campaigns try to reach so-called swing-voters. However, according to the observed re-tweeting behavior, only 3% of all Twitter users re-tweet the content of different political parties. Therefore, the focus on hardly detectable swing-voters might have limited effects.
- The dataset can be used to investigate the limitations in targeting specific citizens. According to this dataset, most observed Twitter users seem to be inactive, making it hard to derive account-specific conclusions. Therefore, the question arises whether micro-targeting approaches are really useful for most social media users.

This dataset might contain some answers for some of these questions. However, this list should be understood as a guiding proposal of what could be done with this dataset. It should not limit any analytical directions or further research ideas. To enable a maximum of use cases, the data are provided as Twitter raw data following the JSON formats defined by Twitter (see Appendix B for an example Tweet)

- Tweet data: <https://dev.twitter.com/overview/api/tweets>
- User data: <https://dev.twitter.com/overview/api/users>
- Entity data (hashtags, media URLs, user mentions): <https://dev.twitter.com/overview/api/entities>

The raw data format can be used to develop specialized analysis tools to extract relevant data from these JSON raw data files.

4.5. Processing the Dataset Using Twista

However, with this dataset, a command line tool suite and Python API called Twista is provided. This shall make analysis of collected Twitter data more straightforward [20]. Twista is provided as a Python package. It can be installed using the following command line instructions.

Listing 1: Installing Twista

```
git clone https://github.com/nkratzke/twista.git
cd twista
pip3 install .
```

5. Dataset Availability

The dataset is provided via Zenodo <https://zenodo.org/deposit/835735> and can be processed using the Twista command line tool [20], which is provided and referenced via Zenodo, as well (<https://doi.org/10.5281/zenodo.845857>). Zenodo [21] is an Open Data platform operated by CERN and initiated by the OpenAIRE European Union research project.

6. Conclusions

This sample dataset was recorded and provided to enable Twitter dataset analysis of the 2017 election campaigns for the 19th German Bundestag. It supplements similar datasets and studies [3,7] for other countries of the European Union to provide a broader picture of social media adaption in political election campaigns. The dataset and this dataset description have been published in order to enable further analytical directions and further research ideas for academic researchers, politicians and election campaign strategists in the political science and political communication research community. However, analysts have to consider the limitations of this dataset; especially that this dataset is not suitable to derive something like party proximity for a specific Twitter account. On the one hand, this is not allowed by the Twitter User Protection terms. However, what is more severe, the provided grouping of users was done by analyzing the re-tweeting behavior only. Additionally, for a large amount of users, only one or two re-tweets were recorded, which is obviously less information to speculate about a party proximity of a specific account. However, taking all of the 120,000 observed accounts and more than 1,200,000 recorded Twitter interactions, analysts can derive worthwhile insights into party-specific utilization of Twitter.

A first analysis showed interesting relations. Almost half of all recorded tweets were re-tweets; a third were replies; a tenth were quotes; and only 5% were status posts. This percentage stayed more or less constant across the complete time frame of observation. Significant changes in these ratios correlated with noteworthy events like the TV debates around 3–5 September, the Barcelona terror attacks on 17 August 2017, the G20 riots in Hamburg around 6 July 2017, the surprising German Bundestag ballot on same-sex marriage on 30 June 2017, and so on.

The analysis of the re-tweeting behavior enables one to group observed users according to political parties without sophisticated and complex natural language processing toolkits like NLTK [18]. Obviously, for other Twitter interactions like quoting or replying, more sophisticated analysis methods are necessary. However, we can draw the conclusion for half of all observed Twitter accounts just by looking at the re-tweeting.

Furthermore, it seems that re-tweeters of populist parties seem to dominate the reply space. This space might contain answers to understand the mechanics of “hate-speech” and maybe the so-called “angry white men”. Additionally, re-tweeters of populist parties seem to dominate quoting interactions, as well. It would be interesting to understand how this is used by populist parties to discredit other political positions. Re-tweets are essential to every party and are the main instrument

to disseminate one's own content. However, re-tweets are applied disproportionately high by populist parties and the group of "unknowns". Additionally, left-wing parties seem to have slightly more re-tweets than conservative or liberal parties. However, more re-tweets do not automatically result in more votes for the respective parties. Therefore, Twitter datasets are no crystal ball for democratic elections [7,9,13], but they might be helpful to understand democratic election results better. This might be especially true for groups who are feeling politically penalized and misunderstood.

Acknowledgments: This research would not have been possible without the general support of the Lübeck University of Applied Sciences and its will to enable arbitrary research according to the freedom of research rights of the German constitution; a right that is trampled too often in many other countries and political systems and, therefore, a right that has to be defended every day.

Conflicts of Interest: The author declares no conflict of interest. He is no member of any of the mentioned political parties or any political party at all. According to the performed grouping of this dataset, his Twitter account @NaneKratzke would be likely rated as an 'inconsistently' re-tweeting account. As a self-assessment, he would classify himself as a swing-voter.

Abbreviations

The following abbreviations are used in this manuscript:

API	Application Programming Interface
JSON	JavaScript Object Notation (a text-based serialization format)
URL	Uniform Resource Locator
U.S.	United States (of America)
TV	Television
ZIP	An archive file format that supports lossless data compression

Appendix A. Followed and Categorized Twitter Screen Names (follow.json)

Listing 2 lists all Twitter screen names of followed politicians. Party categorization is done via the keys of the JSON format. This format is used by Twista [20] for initial tagging of nodes and follow-up tag propagation along re-tweeting edges of the graph.

Listing 2: follow.json

```

"CDU/CSU": [
  "KLimke_CDU", "steffenbilger", "VolkmarKlein", "kretsc", "PSchnieder", "MatthiasHauer", "UweSchummer",
  "IngbertLiebing", "armin_schuster", "plengsfeld", "mechthildheil", "marcusweinberg", "drthomasfeist",
  "HGundelach", "fuchtel", "cducusbt", "ManderlaGisela", "SylviaPantel", "HBraun", "georgnuesslein",
  "rbrinkhaus", "Wellenreuther", "Axel_Fischer", "YvonneMagwas", "hoffmannmdb", "JohannesSingham",
  "AWidmannMauz", "JM_Luczak", "berndfabritius", "amattfeldt", "DrAndreasNick", "karstenmoering",
  "OstermannMdB", "olavgutting", "christianhirte", "SibyllePfeiffer", "jensspahn", "tj_tweets",
  "DoroBaer", "HHirte", "peteraltmaier", "SteinekeCDU", "MaikBeermann", "MGrosseBroemer",
  "StefanKaufmann", "juergenhardt", "charlesmhuber49", "berndsiebert", "meister_schafft",
  "Thomas_Bareiss", "petertauber", "kudlaleipzig", "franksteffel", "tschipanski", "DWoehrl",
  "AlexanderRadwan", "groehe", "AndreaLindholz", "MarkHauptmann", "frankheinrich", "smuellermdb",
  "matthiaszimmer", "julia_obermeier", "dieAlbsteigerin", "schroeder_k", "VolkerUllrich", "koschyk",
  "erwin_rueddel", "Stettenchris", "guenterkrings", "janmetzler", "Manfredbehrens", "stephanharbarth",
  "BettinaHornhues", "fj_josef", "SvenVolmering", "HPFriedrichCSU", "TinaSchwarzer", "KLeikert",
  "marlenemortler", "GudrunZollner", "ruedigerkruse", "thomasgebhart", "RKiesewetter", "kaiwegner",
  "XaverJung", "helmut_nowak", "drmfuchs", "anjaweisgerber", "josteiniger", "eckhardtrehberg",
  "wanderwitz", "NadineSchoen", "jenskoepen", "PeterWeissMdB", "manfred_grund", "MatthiasHeider",
  "hahnflo", "bernhardkaster", "DerLenzMdB", "jungfj", "ninawarken", "RonjaSchmitt", "PatrickSensburg",
  "Kai_Whittaker"
],
"SPD": [
  "MartinSchulz", "sebast_hartmann", "GabiKatzmarek", "thomashitschler", "EskenSaskia", "michaelaengel",
  "SPDUesseldorf", "UlrichKelber", "GabiWeberSPD", "KarambaDiaby", "baerbelbas", "ZieglerMdB",
  "MatthiasIlgen", "kerstin_tack", "AnnetteSawade", "dieschmidt", "josip_juratovic", "michael_thews",
  "DirkWiese4", "PERnstberger", "larscastellucci", "lischkab", "KerstinGriese", "karl_lauterbach",
  "FrankeEdgar", "MartinRabanus", "arnoklare", "Schwarz_MdB", "GabiHillerOhm", "Elke_Ferner",
  "rainerarnold", "soerenbartol", "CanselK", "ulifreese", "zierke", "evahoegl", "MechthildRawert",
  "KaczmarekOliver", "marcobuelow", "MetinHakverdi", "swenschulz", "hubertus_heil", "MartinRosemann",
  "MiRo_SPD", "g_reichenbach", "FrankSchwabe", "BetMueller", "UlliNissen", "larsklingbeil",
  "waltraud_wolff", "SClenne", "achim_p", "A_Gloekner", "DennisRohde", "HildeMattheis", "utevogt",
  "ChristianFlisek", "kahrs", "RebmannMdB", "edrossmann", "chstraesser", "danielakolbe", "JensZimmermann1",

```

```

"soenkerix", "CPetryMdB", "BurkertMartin", "HellmichMdB", "lothar_binding", "matthiasbartke", "oetzoguz",
"FlorianPost", "brigittezypris", "juergencosse", "MarcusHeld_SPD", "NielsAnnen", "florianpronold",
"HiltrudLotze", "michaelgrossmdb", "schneidercar", "rischwasu", "LangeMdB", "muellerchemnitz",
"jakobmierscheid", "ThomasOppermann", "SpinrathNorbert", "W_Priesmeier"
],
"Gruene": [
"nouripour", "TabeaRoessner", "katdro", "katjadoerner", "PeterMeiwald", "KoenigsGruen", "agnieszka_mdb",
"stephankuehn", "F0stendorff", "KonstantinNotz", "RenateKuenast", "MariaKlSchmeink", "IreneMihalic",
"ekindeligoez", "jtrittin", "oetzcanmutlu", "KaiGehring", "Luise_Amtsberg", "steffilemke", "MarkusTressel",
"tobiaslindner", "GrueneBundestag", "fbrantner", "ChrisKuehn_mdb", "W_SK", "SteffiLemke", "GrueneBeate",
"markuskurthmdb", "OetzcanMutlu", "ulle_schauws", "ManuelSarrazin", "beatewaro", "terpeundteam",
"petermeiwald", "GoeringEckardt", "kerstinandreae", "MarieluiseBeck", "Uwekekeritz", "BrigittePothmer",
"Volker_Beck", "GruenSprecher", "ABAerbock", "gruenebundestag", "DorisWagner_MdB", "BriHasselmann",
"die-gruenen", "ebner_sha", "monikalazar", "DieschbourgC", "NicoleMaisch", "renatekuenast", "cem_oezdemir",
"DJanecek", "LisaPaus", "WilmsVal", "sven_kindler", "BaerbelHoehn", "julia_verlinden", "BabettesChefin",
"crueffer", "Oliver_Krischer", "mdb_stroebele"
],
"Linke": [
"karinbinder", "Linksfraktion", "SevimDagdelen", "GUENGL", "Petra_Sitte_MdB", "DietmarBartsch",
"HerbertBehrens", "Diether_Dehtm", "ch_buchholz", "jankortemdb", "NicoleGohlke", "dielinke",
"SuzaKarawanskij", "MWBirkwald", "ernst_klaus", "TeamPetraPau", "WolfgangGehrcke", "rosaluxstiftung",
"AndrejHunko", "GregorGysi", "UllaJelpke", "HeikeHaensel", "WolfgangGehrcke", "RosemarieHein",
"Annette_Groth", "berlinliebich", "jankortemdb", "Team_GLoetzsch", "JuttaKrellmann", "alexandersneu",
"katjakipping", "conni_moehring", "NordMdB", "thlutze", "sabineleidig", "norbert_mdb", "SuzaKarawanskij",
"MichaelLeutert", "niemamovassat", "frank_tempel", "HPetzold", "KirstenTackmann", "axeltroost",
"Petra_Sitte_MdB", "PetraPauMaHe", "SWagenknecht", "katrin_werner", "haraldweinberg", "voglerk",
"jwunderlichbt", "europeanleft", "NicoleGohlke", "CarenLay", "martinarenner", "halina_waw"
],
"FDP": [
"solms", "LFLindemann", "kielclaas", "christianduerr", "MarcoBuschmann", "Lambsdorff", "franksitta",
"fdp", "HaukeHilz", "EUTheurer", "KatjaSuding", "michael_g_link", "dfoest", "c_lindner", "MAstrackZi",
"Wissing", "johannesvogel", "kcortez66740", "MarcusFaber", "Heiner_Garg", "KemmerichThL", "Otto_Fricke",
"KonstantinKuhle", "LindaTeuteberg", "JoachimStamp", "PascalKober", "KH_Paque", "ruppert_stefan",
"jimmyschulz", "ManuelHoferlin", "MarcelKlingeVS", "hansjoachimotto", "HAHNmeint", "JPirscher",
"FlorianOtt", "Stefan_Birkner", "BundesLHG", "LoeningMarkus", "FDPEuropa"
],
"AfD": [
"poggenburgandre", "AfdKompakt", "arminpaulhampel", "fraukepetry", "joerg_meuthen",
"beatrix_vstorch", "georg_pazderski", "julianflak", "marcuspretzell", "AfdKompakt", "afd_bund",
"TrauDichDE"
]
]

```

Appendix B. Example Tweet (Raw Data)

The following code snippet presents an example tweet encoded in JSON. Some data are shortened (text and user description), and some data are omitted (visual profile data) for better readability.

Listing 3: Example tweet (JSON format)

```

"created_at": "Thu Jun 01 20:20:00 +0000 2017",
"id": 870374393628827648,
"id_str": "870374393628827648",
"text": "Unsere Gr\u00f6\u00dfte Antwort auf Trumps Entscheidung: Raus aus der Kohle #Kohleausstieg [...]",
"source": "<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client</a>",
"truncated": false,
"in_reply_to_status_id": null,
"in_reply_to_status_id_str": null,
"in_reply_to_user_id": null,
"in_reply_to_user_id_str": null,
"in_reply_to_screen_name": null,
"user": {
  "id": 4110374301,
  "id_str": "4110374301",
  "name": "Gerhard Schick",
  "screen_name": "SchickGerhard",
  "location": "Mannheim & Berlin",
  "url": "http://www.gerhardschick.net",
  "description": "Mitglied des Deutschen Bundestages, finanzpolitischer Sprecher bei @GrueneBundestag [...]",
  "protected": false,
  "verified": true,
  "followers_count": 4762,
  "friends_count": 176,
  "listed_count": 74,

```

```

    "favourites_count": 49,
    "statuses_count": 2373,
    "created_at": "Wed Nov 04 07:36:26 +0000 2015",
    "utc_offset": null,
    "time_zone": null,
    "geo_enabled": false,
    "lang": "de",
    "contributors_enabled": false,
    "is_translator": false,
    "following": null,
    "follow_request_sent": null,
    "notifications": null
  },
  "geo": null,
  "coordinates": null,
  "place": null,
  "contributors": null,
  "is_quote_status": false,
  "retweet_count": 0,
  "favorite_count": 0,
  "entities": {
    "hashtags": [
      { "text": "Kohleausstieg", "indices": [65, 79] },
      { "text": "Divestment", "indices": [86, 97] },
      { "text": "Klimaschutzabkommen", "indices": [99, 119] }
    ],
    "urls": [
      {
        "url": "https://t.co/0eJ00M3kSH",
        "expanded_url": "https://dbtg.tv/fvid/7115264",
        "display_url": "dbtg.tv/fvid/7115264",
        "indices": [120, 143]
      }
    ],
    "user_mentions": [],
    "symbols": []
  },
  "favorited": false,
  "retweeted": false,
  "possibly_sensitive": false,
  "filter_level": "low",
  "lang": "de",
  "timestamp_ms": "1496348400872"

```

References

1. Issenberg, S. How Obama's Team Used Big Data to Rally Voters. MIT Technology Review, 19 December 2012. Available online: <https://www.technologyreview.com/s/509026/how-obamas-team-used-big-data-to-rally-voters/> (accessed on 19 October 2017).
2. Wagner, J. 'Clinton's Data-Driven Campaign Relied Heavily on an Algorithm Named Ada. What Didn't She See? Washington Post, 9 November 2016. Available online: https://www.washingtonpost.com/news/post-politics/wp/2016/11/09/clintons-data-driven-campaign-relied-heavily-on-an-algorithm-named-ada-what-didnt-she-see/?utm_term=.7f86c9d90768 (accessed on 19 October 2017).
3. Barberá, P.; Rivero, G. Understanding the Political Representativeness of Twitter Users. *Soc. Sci. Comput. Rev.* **2015**, *33*, 712–729.
4. 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. *CoRR* **2013**, *arXiv:1306.5204*.
5. Wang, Y.; Callan, J.; Zheng, B. Should We Use the Sample? Analyzing Datasets Sampled from Twitter's Stream API. *ACM Trans. Web* **2015**, *9*, 1–23.
6. Abreu, J.; Almeida, P.; Silva, T. From Live TV Events to Twitter Status Updates—A Study on Delays. In *Applications and Usability of Interactive TV*; Springer International Publishing: Cham, Switzerland, 2016; Volume 605, pp. 105–120.
7. Jungherr, A. Tweets and Votes, a Special Relationship: The 2009 Federal Election in Germany. In Proceedings of the 2nd Workshop on Politics, Elections and Data, PLEAD '13, San Francisco, CA, USA, 28 October 2013; pp. 5–14.

8. Wang, H.; Can, D.; Kazemzadeh, A.; Bar, F.; Narayanan, S.S. A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. In Proceedings of the ACL 2012 System Demonstrations, ACL'12, Jeju Island, Korea, 10 July 2012.
9. Gayo-Avello, D.; Metaxas, P.T.; Mustafaraj, E. Limits of Electoral Predictions Using Twitter. In Proceedings of the 5th International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
10. Straus, J.R.; Glassman, M.E. *Social Media in Congress: The Impact of Electronic Media on Member Communications Analyst on the Congress*; Technical Report; Congressional Research Service: Washington, DC, USA, 2016.
11. Cook, J.M. Twitter Adoption and Activity in U.S. Legislatures: A 50-State Study. *Am. Behav. Sci.* **2017**, *61*, 724–740, doi:10.1177/0002764217717564.
12. Waisbord, S.; Amado, A. Populist communication by digital means: Presidential Twitter in Latin America. *Inf. Commun. Soc.* **2017**, *20*, 1330–1346, doi:10.1080/1369118X.2017.1328521.
13. Sang, E.T.K.; Bos, J. Predicting the 2011 Dutch Senate Election Results with Twitter. In Proceedings of the Workshop on Semantic Analysis in Social Media, Avignon, France, 23 April 2012.
14. Tumasjan, A.; Sprenger, T.; Sandner, P.; Welp, I. Predicting elections with twitter: What 140 characters reveal about political sentiment. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, USA, 23–26 May 2010; pp. 178–185.
15. Papakyriakopoulos, O.; Shahrezayeh, M.; Thielges, A.; Medina Serrano, J.C.; Hegelich, S. Social Media und Microtargeting in Deutschland. *Inform. Spektrum* **2017**, *40*, 327–335.
16. Speriosu, M.; Sudan, N.; Upadhyay, S.; Baldridge, J. Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph. In Proceedings of the 1st Workshop on Unsupervised Learning in NLP, EMNLP'11, Edinburgh, Scotland, 30 July 2011; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; pp. 53–63.
17. Fraiser, O.; Cabanac, G.; Pitarch, Y.; Besançon, R.; Boughanem, M. Uncovering Like-minded Political Communities on Twitter. In Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2017, Amsterdam, The Netherlands, 1–4 October 2017; pp. 261–264.
18. Bird, S.; Klein, E.; Loper, E. In *Natural Language Processing with Python*, 1st ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2009.
19. Hagberg, A.A.; Schult, D.A.; Swart, P.J. Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA, USA, 19–24 August 2008; pp. 11–15.
20. Kratzke, N. Twista—A Twitter Streaming and Analysis Tool Suite, 2017. Available online: <https://doi.org/10.5281/zenodo.845857> (accessed on 19 October 2017).
21. Manola, N.; Rettberg, N.; Manghi, P. OpenAIREplus Project Executive Report. Technical Report, 2015. Available online: <https://doi.org/10.5281/zenodo.15464> (accessed on 19 October 2017).



© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).