SSOAR

Open Access Repository
www.ssoar.info

Systematically Monitoring Social Media: the case of the German federal election 2017

Stier, Sebastian; Bleier, Arnim; Bonart, Malte; Mörsheim, Fabian; Bohlouli, Mahdi; Nizhegorodov, Margarita; Posch, Lisa; Maier, Jürgen; Rothmund, Tobias; Staab, Steffen

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Stier, S., Bleier, A., Bonart, M., Mörsheim, F., Bohlouli, M., Nizhegorodov, M., ... Staab, S. (2018). *Systematically Monitoring Social Media: the case of the German federal election 2017.* (GESIS Papers, 2018/04). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.21241/ssoar.56149

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by-nc/4.0/deed.de

Terms of use:
This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more Information see:
https://creativecommons.org/licenses/by-nc/4.0

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-56149-4

gesis Leibniz-Institut für Sozialwissenschaften

GESIS Papers

2018|04

Systematically Monitoring Social Media:

The case of the German federal election 2017

Sebastian Stier, Arnim Bleier, Malte Bonart, Fabian Mörsheim, Mahdi Bohlouli, Margarita Nizhegorodov, Lisa Posch, Jürgen Maier, Tobias Rothmund & Steffen Staab

GESIS Papers 2018|**04**

Systematically Monitoring Social Media:

The case of the German federal election 2017

Sebastian Stier[1], *Arnim Bleier*[1], *Malte Bonart*[1], *Fabian Mörsheim*[2], *Mahdi Bohlouli*[2], *Margarita Nizhegorodov*[1], *Lisa Posch*[1], *Jürgen Maier*[2], *Tobias Rothmund*[2] *& Steffen Staab*[2]

[1] GESIS – Leibniz Institute for the Social Sciences
[2] University of Koblenz · Landau

GESIS – Leibniz-Institut für Sozialwissenschaften 2018

GESIS Papers

GESIS – Leibniz Institute for the Social Sciences
Computational Social Science
Social Analytics and Services
Unter Sachsenhausen 6-8
50667 Köln
Telefon: 0221 / 476 94 - 221
Telefax: 0221 / 476 94 - 199
E-Mail: sebastian.stier@gesis.org

ISSN: 2364-3781 (Online)
Herausgeber,
Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

Contents

1 Introduction 5
2 Related Work 6
3 Defining the Political Communication Space 7
3.1 Candidates 8
3.2 Organizations: Political Parties and Gatekeepers 9
3.3 Political Topics 9
4 Data Collection 10
4.1 Data Retrieval from Twitter 11
4.2 Data Retrieval from Facebook 11
5 Exploratory Analysis 12
5.1 Data Cleaning 12
5.2 Temporal Patterns 13
5.3 Attention on Twitter 14
5.4 Activity and Resonance by Party 15
6 Data Sharing 18
7 Conclusions 19
Appendices 24
A Appendix 24

1 Introduction

The campaign for the German parliament (*Bundestag*) 2017 was finally the one in which party strategists and observers regarded social media not just as experimental and in essence peripheral venues, but rather as important arenas where elections are won or lost. The year before, Donald Trump won the presidency in the U.S., which was attributed to his authentic Twitter use and a skillful mobilization of supporters via social media. Right-wing populist forces on the rise in Germany like the AfD or Pegida similarly use social media to bypass media gatekeepers and reach sympathetic target audiences (Stier, Posch, Bleier, & Strohmaier, 2017). So-called "fake news", social bots (semi-automated accounts) and online propaganda (e.g., orchestrated by Russia), accompanied by a growing mistrust of legacy media in the wake of the refugee crisis threatened to impact the campaign. The agenda-setting power of online media (Russell Neuman, Guggenheim, Mo Jang, & Bae, 2014) became apparent when an online campaign led by party activists helped to propel Martin Schulz and the SPD to a parity with the CDU in public opinion polls in March 2017 ("Schulzzug"). In the campaigning arena, parties applied innovations like micro-targeting at a larger scale in order to harness the persuasive potential of social media. Importantly, the intense – maybe even over-proportional – coverage of the aforementioned phenomena by the mass media contributed to the perception of an increased political role of social media.

In order to understand these processes and also to keep economic actors like Facebook and Twitter accountable for their political influence, it is essential for academia to systematically explore new digital data sources. Yet, online social networks are complex and intransparent sociotechnical web environments (Strohmaier & Wagner, 2014). Therefore, it is a considerable task to collect digital trace data at a large scale and at the same time adhere to established academic standards. In the context of political communication, important challenges are (1) defining the social media accounts and posts relevant to the campaign (*content validity*), (2) operationalizing the venues where relevant social media activity takes place (*construct validity*), (3) capturing all of the relevant social media activity (*reliability*), and (4) sharing as much data as possible for reuse and replication (*objectivity*).

The present collaborative project by GESIS – Leibniz Institute for the Social Sciences and the E-Democracy Program of the University of Koblenz-Landau conducted such an effort. We concentrated on the two social media networks of most political relevance, Facebook and Twitter. These platforms have different architectures, user bases and usage conventions that need to be taken into account conceptually and methodologically. Section 2 discusses previous work related to our endeavor. In Section 3, we lay out what kinds of activities we define as part of the "political communication space" on Facebook and Twitter. Section 4 outlines our data collection. In Section 5, we present exploratory findings of how political communication on the Bundestag campaign unfolded on social media. Section 6 discusses how we share the main output of the project, the "BTW17 dataset" and lists of election candidates and their social media accounts. We conclude with a critical reflection of our efforts and an outline of future research avenues in Section 7.

2 Related Work

A burgeoning literature has analyzed the role of social media during election campaigns. In this section, we first structure related work according to the logic of data collection. Then we summarize methodological research that has revealed several pitfalls when collecting social media data. Finally, we discuss related work on the Bundestag campaign 2017. The predominant focus in the existing literature clearly lies on Twitter use during election campaigns.[1] Generally, one can distinguish between audience-centered and elite-centered data collections.

Audience-centered designs capture tweets containing a set of hashtags ("#btw17") or keywords related to a campaign ("cdu", "merkel"). The goal is to reveal how interested users participate in political debates. From this body of research, we have a pretty good understanding of the dynamics that unfold during election campaigns in the public Twitterverse. A particular focus lies on "second screening" during TV debates (Freelon & Karpf, 2015; Lin, Keegan, Margolin, & Lazer, 2014; Trilling, 2015; Vaccari, Chadwick, & O'Loughlin, 2015), influential users (Dubois & Gaffney, 2014; Freelon & Karpf, 2015; Jürgens, Jungherr, & Schoen, 2011), attempts to predict election results (DiGrazia, McKelvey, Bollen, & Rojas, 2013; Tumasjan, Sprenger, Sandner, & Welpe, 2010) and the identification of central political topics (Bruns, Burgess, et al., 2011; Jungherr, Schoen, & Jürgens, 2016; Trilling, 2015).

Elite-centered designs concentrate on a well-defined subset of the Twitterverse, namely specific accounts of interest. In contrast to audience-centered designs focusing on the demand side of politics, elite-centered studies concentrate on the supply side of politics, i.e., politicians, parties or journalists. Moreover, not only the tweets of these users mentioning a political keyword are of interest, but rather their overall behavioral patterns on social media. Here, studies focused on the adoption of platforms by politicians (Quinlan, Gummer, Roßmann, & Wolf, 2017; Vergeer, Hermans, & Sams, 2013), the (partisan) structure of online networks of candidates (Aragón, Kappler, Kaltenbrunner, Laniado, & Volkovich, 2013; Lietz, Wagner, Bleier, & Strohmaier, 2014) and their resonance with audiences (Kovic, Rauchfleisch, Metag, Caspar, & Szenogrady, 2017; Nielsen & Vaccari, 2013; Yang & Kim, 2017).

Studies of Facebook are more sparse, which is likely due to technical and privacy constraints. Most Facebook profiles and communication are private, whereas on Twitter, most posts and user profiles are public. Since retrieving posts via keyword search is only possible for (the few) public posts, studies of election campaigns on Facebook necessarily have to concentrate on users with public pages. Thus, most electoral studies are elite-centered in that they concentrate on profiles of politicians and parties (e.g., Caton, Hall, & Weinhardt, 2015; Kovic et al., 2017; Lev-On & Haleva-Amir, 2016; Nielsen & Vaccari, 2013; Williams & Gulati, 2013). From this preselection, the analysis sometimes still moves to the audience, for example the politically active users in the comments sections of these pages (e.g., Freelon, 2017).

Our project also draws on previous studies that critically assessed the data provided by Twitter's Application Programming Interface (API). Driscoll and Walker (2014) performed a comparison between the publicly accessible Streaming API (providing up to 1% of live Twitter traffic) and the Gnip PowerTrack Firehose dataset provided by Twitter (granting full access to the Twitter data stream). They found that during periods of public contention like election campaigns or protests, the Streaming API returns biased data because the rate limit tends to be surpassed when a social phenomenon generates a lot of public interest. Similarly, Morstatter, Pfeffer, Liu, and Carley (2013) found that as a researcher increases the parameters (keywords) of interest, the coverage of the Streaming API

[1] It is not our goal to review this literature in its entirety. Instead, we focus on commonalities in the predominant data collection strategies. For a comprehensive literature review on Twitter use during election campaigns, see Jungherr (2016).

decreases. In the most systematic evaluation to date, Tromble, Storz, and Stockmann (2017) compared the Streaming API and the Search API to a Firehose dataset as a ground truth. While they identified serious biases in the Search API, they found that the Streaming API returns "nearly all" relevant tweets when rate limits are not reached. Based on this methodological research, we design a data collection scheme that is robust to the biases inherent to the Twitter API (see Section 4).

In contrast to Twitter, the Facebook Graph API allows the collection of all non-deleted posts for an unlimited research period. Thus, researchers can design lists of accounts they want to monitor and even collect data ex post, after an election campaign has unfolded. This has been the standard approach in academic works, which, however, misses activities such as posts, comments and likes that got deleted in the meantime. In a first systematic analysis of deletion rates, Bachl (2017) showed that approximately 18% of user generated content on German political Facebook pages had been deleted over the span of eight months. Among posts by political actors themselves, only 2.3% of posts could not be retrieved anymore.[2] Our own data collection from Facebook generated two independent datasets that are susceptible to these limitations to various degrees (see Section 4).

As for studies related to the current Bundestag campaign, Schmidt (2017) collected social media accounts of all candidates. In his study, he describes the adoption of social media by candidates and their follower relationships on Twitter. Yet, in addition to this metadata that can also be accessed ex post, our project also collected the contents of candidates' tweets as well as candidates' interactions with other users in real time. Another project collected data at a larger scale, albeit exclusively on Twitter (Kratzke, 2017).[3] Apart from being limited to one platform, this project only collected data for 360 politicians, mostly sitting members of the Bundestag. Such a convenience sample only allows for limited substantive analyses.

In the following section, we describe which target concepts we consider essential to be covered during an election campaign. Furthermore, we describe how they can be operationalized on social media. In our definition of the "political communication space", we chose a hybrid approach that integrates both, audience-centered and elite-centered research designs.

3 Defining the Political Communication Space

Our goal was to collect all publicly available political communication related to the Bundestagswahl on Facebook and Twitter. To this end, we define three target concepts on which political communication is typically centered: (1) politicians, here Facebook pages and Twitter accounts of candidates in the election campaign, (2) Facebook pages and Twitter accounts of political parties and gatekeepers such as media organizations, and (3) keywords denoting central political topics on Twitter. This holistic conceptualization not only covers the most important actors, but also aims to capture the online engagement of regular citizens with politics. For politicians and parties, we have confined our data to those parties that had, based on polls, a realistic chance to win seats in parliament.[4] Figure 1 visualizes the conceptualization which we will explain in detail in this section.

[2] We could not incorporate these insights in our project, as Bachl presented first findings in a conference presentation only on 22 September 2017.
[3] We are aware that various other teams collected specific Twitter data for concrete research questions as well.
[4] Included parties are AfD, Bündnis 90/Die Grünen, CDU, CSU, FDP, Linke and SPD.

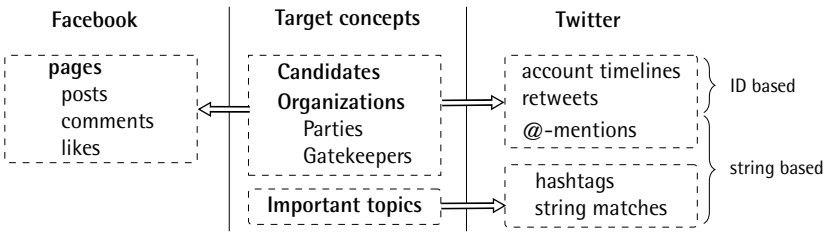


Figure 1: The political communication space and its operationalization.

3.1 Candidates

The official list of candidates eligible for the election was not published by the *Bundeswahlleiter* until mid of August, just a month before election day and therefore right in the middle of the campaign. However, our goal was to cover campaign activity as early as possible. Hence, we had to devote considerable efforts to first compile lists of candidates. This process was guided by the specifics of the electoral system for the Bundestag, mixed-member proportional representation. 50 percent of members of parliament are elected via relative majority in electoral districts (direct candidates). The other 50 percent are chosen from party lists elected by the regional divisions of parties in the 16 federal states (list candidates).

In case of list candidates, the process of identification was helped by the fact that most parties at state level published this information on their websites rather early. For the direct candidates, a comprehensive (web) search was required, because the nomination process is based at the local level. The preparation of the candidate list was a continuous process with the benchmark being the official list published in August.[5]

Based on the candidate lists we identified the respective Facebook pages and Twitter accounts of the candidates. The main source were the websites of politicians and parties linking to social media accounts. In other cases we used Facebook's and Twitter's search and look-up functionality along the following (soft) criteria: Do the name, self-description and/or photos correspond to the politician? Is the account mentioned/linked in relevant conversations? Is the content of post real (or satire, e.g., in the case of Angela Merkel parody accounts on Twitter)? For Twitter, we collected one account per person and for Facebook, we collected up to two accounts, regardless of being a page or profile.

After the election, we compared our data to two similar datasets that became public at the time.[6] While we covered a substantially larger number of accounts (199% more than the Tagesspiegel, 195% more than the Open Knowledge Foundation), we still added 151 Facebook accounts and 52 Twitter accounts from these datasets to our final list.

The data on candidates including available attributes and their social media accounts can be downloaded from Stier, Bleier, Bonart, et al. (2018).

[5] Our original search missed only two percent of candidates (52 of the 2,516 final candidates). On the other hand, we identified 45 individuals who did not end up being an official candidate. The latter resulted mostly from disorganized/fluid processes and conflicts within party organizations, especially in the case of the AfD.

[6] These are a collection made by the Tagesspiegel featuring candidates' Facebook and Twitter accounts (http://wahl.tagesspiegel.de/2017/kandidatenbank) and data from the Open Knowledge Foundation on Facebook accounts (https://docs.google.com/spreadsheets/d/1l1JO13aDCcTFHoVD0KjWFHSCe6h3joPjBiSJxKcZFkE).

3.2 Organizations: Political Parties and Gatekeepers

Besides candidates, there are additional influential accounts on social media during election campaigns, most importantly, accounts by political parties and news media. Likewise, their Twitter accounts and Facebook pages were researched for the construction of an "organizations" dataset.

The list of relevant party actors contains all accounts of the national parties, their caucuses in the Bundestag ("Bundestagsfraktion"), the parties in the federal states ("Landesparteien") and youth organizations (e.g., the Jusos). Parties at all these different levels predominantly focused on the Bundestag campaign during our research period. At the same time, future research should also incorporate the local party branches (e.g., CDU Cologne).

In addition to these party actors, we also compiled lists of accounts belonging to the right-wing protest movement Pegida that generate a lot of engagement on German social media (Stier et al., 2017). It was to be expected that these accounts would form part of the right-wing media ecosystem during the election campaign.

In order to compile a list of media accounts, we crowdsourced the account names of German media present on Facebook and Twitter. On the crowdsourcing platform CrowdFlower[7], we asked German crowdworkers to find links to German media accounts ("mainstream media" as well as "alternative media") that report on political topics. The rationale behind using (and paying) crowdworkers for this task was to construct a diverse set of accounts that also captures the long tail of media accounts. In total, we collected 6,211 responses from crowdworkers on Twitter media accounts (2,815 unique accounts) and 4,774 responses on Facebook media accounts (2,049 unique accounts). All media accounts with more than one mention were then manually looked through by the first author who screened out non-German accounts and non-related genres like YouTube Stars. The final list of media pages on Facebook contains 285 accounts, the list of media accounts on Twitter comprises 310 accounts.

3.3 Political Topics

Finally, to monitor central political topics we built a list of related keywords ("selectors"). We predefined a broad set of relevant keywords before the election and decided not to add additional selectors during the campaign. The main reason for this decision is that if a researcher includes new trends, she will always be too late as these can only be identified when they are already popular on social media or in public debates. Thus we first opted for the more systematic approach with a fixed set of topics, before retrospectively capturing additional campaign topics, as explained later.

The selection of keywords related to a topic is a process prone to human biases (King, Lam, & Roberts, 2017). Thus, our choice of selectors is certainly not objective but rather reflects ex ante expectations by the project team which topics and actors would become important during the campaign. We also had to consider two potentially distorting aspects. First, some keywords were used very frequently in other languages and would have flooded our data collection. For instance, we found out that "fdp" is an abbreviation very frequently used in French and Portuguese tweets. Thus we added some keywords with a hashtag, e.g. "#fdp" that most often refers to the German party. Second, some surnames are not unique identifiers for German politicians. The word "gabriel", e.g., is a word used very often in Spanish, thus we included the full name "sigmar gabriel".

We minimize the number of missing relevant messages as we have such a holistic conceptualization of political actors and topics. Many messages on newly emerging topics still contain mentions of

[7] www.crowdflower.com

known candidates, party names or leading candidates. However, the share of relevant communication neglected by our definition of the political communication space is impossible to evaluate – a bias that is inherent to any such data collection effort.

As a starting point for the construction of our selectors list, we used the keywords already collected in the social media data collection conducted by GESIS in 2013 (Kaczmirek et al., 2014). Many of these are universally relevant (e.g., "finanzpolitik" or "tvduell"). Furthermore, collecting these again enables us to directly compare two federal election campaigns in Germany.

Analogous to 2013, we added keywords referring to political institutions and democracy in general, i.e., the German *polity* (Appendix A1). Based on the list from 2013 and an investigation of the recent political agenda, we added keywords on issues (*policy*). This includes traditional fields such as social policy or economic policy, but also topics of particular importance in 2017. For the list of selectors related to policy, we refer to Appendix A2.

We also captured keywords related to *politics*, in particular on the ongoing election campaign (Appendix A3). Among these are popular generic hashtags like "btw17" and political TV events ("tvduell") that typically generate a lot of attention on Twitter (Freelon & Karpf, 2015; Jungherr et al., 2016; Stier, Bleier, Lietz, & Strohmaier, 2018; Trilling, 2015).

Last, our selectors list comprised keywords related to political parties (Appendix A4). This includes the names of the established parties that are not only @-mentioned as organizational accounts very frequently, but also mentioned in free text (as "political topics" in our logic) without an @-prefix. Moreover, we include the names of the leading candidates ("Spitzenkandidaten"). This is of particular importance as the leading candidate of the CDU Angela Merkel does not have a Twitter account. We also included the names of cabinet members in the federal government and the prime ministers of German states ("Bundesländer"), as of beginning of July 2017.

During the campaign, we assembled keywords that were relevant but not captured by our original list.[8] In order to capture these ex post, we used a script to scrape tweets containing these keywords from the Twitter interface allowing us to track topics until their first emergence. Yet this approach has the limitation that retweets are excluded from the interface and that deleted tweets are not available anymore. We were able to capture up to 100 retweets of these original tweets by retrospectively connecting to the Twitter REST API. Through this process, we added additional 2,136,620 tweets to our data. While this approach is not as systematic as the live streaming and has several inherent biases, we opted for still adding these tweets to our collection. We will evaluate the performance of this procedure with data we are currently buying from the data vendor Twitter Enterprise.

4 Data Collection

At the outset, it is important to note that we only collect publicly available information from social media. We collect these digital traces of user activity from the Facebook Graph API and the Twitter Streaming API. In principal, all online interactions of users such as politicians and politically active individuals can be relevant for social science research. Building on the above defined political communication space, we are first interested in the tweets and Facebook posts of political candidates and organizations. Second, we collect the engagement of users with these contents –

[8] These included "weidel", the AfD politician who only rose to prominence during the campaign, "flüchtling" and "migranten" which we missed in our original list, and the party campaign hashtags "fedidwgugl", "traudichdeutschland", "holdirdeinlandzurück", "denkenwirneu", "lustauflinks", "darumgruen", "zeitfürmartin".

retweets and @-mentions on Twitter, comments, shares and likes on Facebook. Third, we likewise consider keywords in messages. Finally, we also collect data from Twitter on follower relationships. The following subsections elaborate on how we technically monitored this political communication space.

4.1 Data Retrieval from Twitter

While Twitter can provide a transparent source of data, achieving this transparency requires to be precise on the data collection process. Building on the set of account names as described above, we used Twitter's Streaming API[9] to collect messages sent by candidates, replies and retweets of these messages as well as messages sent to candidates (@-mentions), starting on 5 July. Using the same method, we collected the analogous messages for political parties and gatekeepers (the selection logic "organizations" described before). Furthermore, we captured messages containing at least one keyword from our list of political topics. The employed software includes Tweepy[10] and Twitter4J[11] for retrieval as well as MongoDB[12] for storing the data. Overall, we took extraordinary precautions to ensure the completeness of the collected data such as independent parallel collections by the project teams at our two institutions.

Furthermore, we retrieved the follower graph for the monitored candidate accounts on 27 June 2017, 14 September 2017, 1 November 2017 and 9 January 2018. That way we know who followed politicians and whom they followed at different time points during the campaign. We retrieved the much larger follower graph for party and organizations accounts on 1 July 2017 (collecting 48 million connections).

4.2 Data Retrieval from Facebook

As elaborated above, the data that can be accessed and monitored publicly differs between social media platforms. On Facebook, a researcher can only collect data from public pages. As a consequence, political posts from non-public accounts (that many candidates also used) and conversations among Facebook friends or in closed groups can not be analyzed. The public portion of Facebook data can be accessed via the Facebook Graph API.[13]

Given these constraints, we define the political communication space on Facebook as posts, comments and likes on public political pages by candidates and organizations. We set up two data collections. The first scheme collected data using an eight-day rolling time window starting on 15 August. This means that comments and likes on a given post were collected eight days after its creation. The rationale behind this approach is that most activity happens in the few days after a post had been created. We also set up a second dataset by retrieving Facebook data from all target accounts after the campaign. However, we recently got aware of the findings of Bachl (2017) who showed that such ex post data collections suffer from considerable deletion rates. Thus, we opted to report data from the first, continuous collection made during the campaign – even though its temporal coverage is more limited.

[9] http://developer.twitter.com/en/docs/tweets/filter-realtime
[10] http://tweepy.org
[11] http://twitter4j.org
[12] http://mongodb.com
[13] http://developers.facebook.com/docs/graph-api

5 Exploratory Analysis

This section presents several exploratory analyses of our dataset covering the research period from 6 July to 30 September, one week after the election. We first evaluate available language filters that should be applied when analyzing a national election campaign. Next, we explore our dataset, focusing on (1) temporal patterns, (2) attention patterns on Twitter, i.e., which topics were the most salient in public discussions on the campaign and (3) activities and engagement by political parties.

5.1 Data Cleaning

As discussed in Section 3, every conceptualization of the political communication space will undoubtedly include posts unrelated to German politics (e.g., "özdemir" is the name of a leading candidate from the German Green party, but also a popular name in Turkey). This "noise" is less of a problem on Facebook, since all posts in our dataset have been contributed to a Facebook page relevant to German politics. Even if a user comments in another language choosing an entirely non-political topic on a page of one of our target accounts, this user still has intentionally chosen to contribute to the German political communication space.

However, on Twitter, where we take out tweets from a large universe of messages posted in many different languages, ambiguities in the meaning of keywords become more of a problem. Moreover, international debates on German leaders or the campaign are not relevant or even distorting in the context of most social science research questions that still focus on the national arena. Many tweets on Angela Merkel, for instance, are coming from outside of Germany. These are related to her role in world politics, her relationship with Donald Trump or are very critical of her refugee policies (those tweets are often coming from the U.S. alt right).

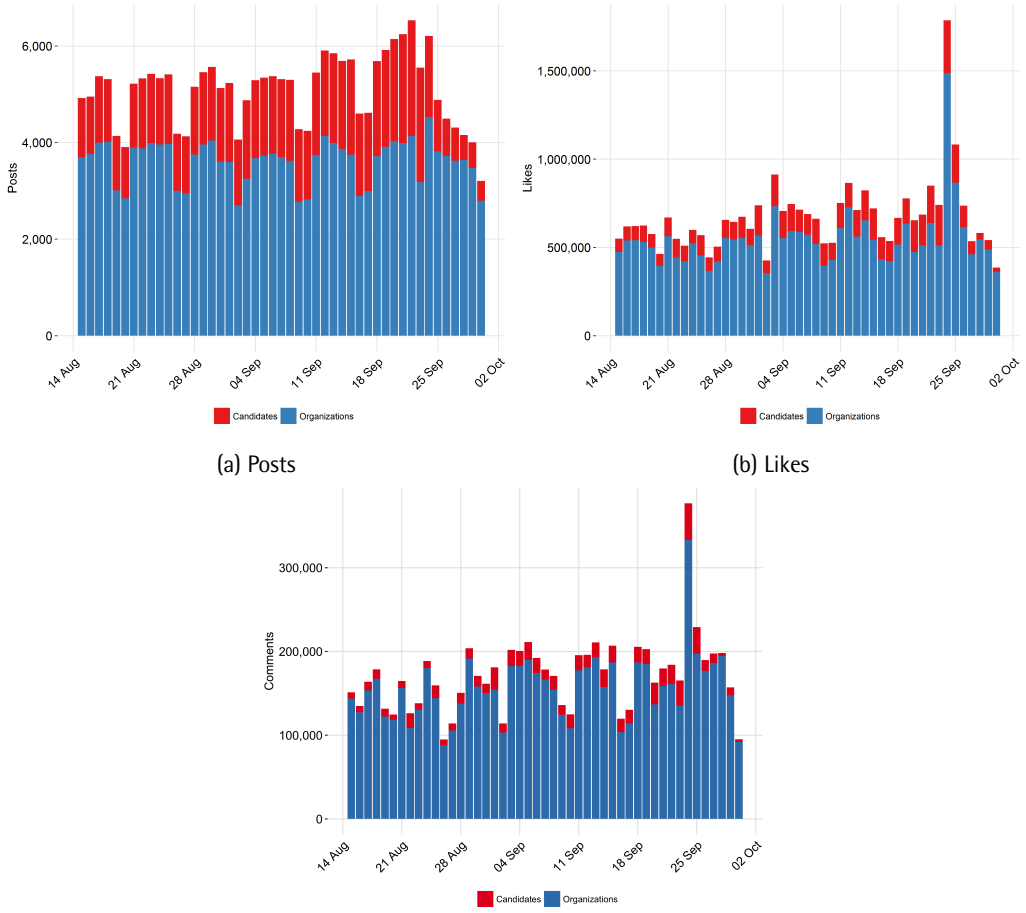
Table 1: Evaluation of language filters

	False positives	**False negatives**
Twitter language interface DE	14 (0.028%)	121 (0.242%)
Twitter machine learning DE	2 (0.004%)	15 (0.03%)

In order to filter out tweets from non-German users and also reduce the amount of topically non-related "noise", we applied and evaluated two filtering approaches.[14] The first approach restricted the dataset to only those tweets by users who chose the interface language German in their Twitter settings (Jungherr et al., 2016). Second, we selected only tweets that were labeled as German by Twitter's machine learning algorithm. Then we randomly sampled and coded 4×500 tweets from our dataset in order to assess the false negative and false positive rates of each approach. Results are reported in Table 1.

It becomes clear that the Twitter machine learning approach is superior to a filtering based on the interface language. A non-trivial percentage of users who are tweeting in German on the election campaign are running their interfaces in other languages – most often English. Meanwhile, the

[14] One additional source of "noise" is non-human activity. Our approach filters out bot activity in languages other than German which targets trending hashtags with spam and advertisements. Moreover, a recent study found that the percentage of messages sent by political bots on Twitter on the German election was only minor (Neudert, Kollanyi, & Howard, 2017). Note, however, that this study was based on an ad hoc sample of one million tweets on only select hashtags. Additionally, the authors define "highly automated accounts" as users tweeting more than 50 times per day on the election, which is certainly debatable.



(a) Posts

(b) Likes

(c) Comments

Figure 2: Facebook activities in our dataset by selection logic.

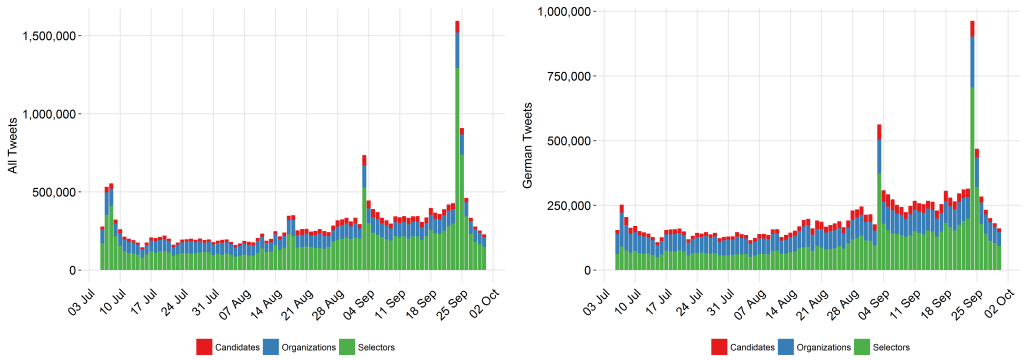
machine learning misclassifies only 2 of 500 non-German tweets as German and 15 of 500 German tweets as non-German. In total, the language of only 0.017% of tweets is classified incorrectly by Twitter's machine learning algorithm. Therefore, we use the tweets that have been labeled as German by Twitter for our exploratory analysis of the Bundestag election campaign.

5.2 Temporal Patterns

We start our presentation of results with a description of the Facebook and Twitter datasets. The first purpose is to describe the raw data we collected before moving to more substantive analyses.

Figure 2 shows the daily Facebook activity over time as measured by posts, comments and likes. As is to be expected, the activity clearly increases with election day on 24 September approaching. On election night, Facebook posts by candidates, party organizations and gatekeepers received more than 1.5 million likes and more than 300,000 comments.

Figure 3 displays the analogous data for Twitter, except that here we were also able to collect tweets



(a) Unfiltered (b) With German filter

Figure 3: Tweets in our dataset by selection logic.

containing keywords on political topics. Panel (a) shows the unfiltered dataset, that is all tweets we collected.[15] Just like in the case of Facebook, activity on Twitter is highest around election day. However, there are four more spikes in the time series: the first two, the G20 summit in Hamburg on 7/8 July 2017 and the terror attack in Barcelona on 17 August 2017, are not immediately related to the Bundestag campaign. Nonetheless, both events were related to political actors and topics that are prominently featured in our political communication space. The third noteworthy peak in Twitter activity is related to the "TV Duell" between the leading candidates Merkel (CDU) and Schulz (SPD) on 3 September 2017.

The figures indicate that Twitter is the more event-driven medium, as the increases in volume related to ongoing political events are usually of a higher magnitude than on Facebook. It is an established finding in political communication that the Twitterverse intensively reacts to TV events during election campaigns. In fact, politicians themselves prefer Twitter over Facebook for "dual screening" purposes (Stier, Bleier, Lietz, & Strohmaier, 2018).

As discussed before, all posts by media organizations were included in these figures based on the raw data. Many of the Facebook posts and tweets by media organizations are of course non-political and thus not immediately relevant to the campaign. We provide the account handles and IDs of these accounts (Stier, Bleier, Bonart, et al., 2018), which allows a researcher who has reconstructed our dataset to remove tweets, Facebook posts and interactions of these accounts if preferred for her analysis.

5.3 Attention on Twitter

This section explores the attention the Twitterverse devoted to different topics and actors. Twitter comes closest to an online "public sphere" that reacts to external events happening during a campaign. In contrast, Facebook is a medium dominated by reciprocal interactions, thus, the page owner has more influence on setting the agenda by choosing which topics to post on.

Figure 4 shows the most frequently mentioned political topics as predefined by the selectors found in Appendix A1 to A4.[16] Most attention was devoted to the AfD, followed by mentions of the two

[15] Note that this figure includes duplicates since a tweet will be collected twice if it addresses multiple target concepts. To give an example including a mention of a candidate and a topic selector: "@MartinSchulz has won the #tvduell".

[16] As we are interested in general attention patterns, we also include the appearance of selectors in combination with other

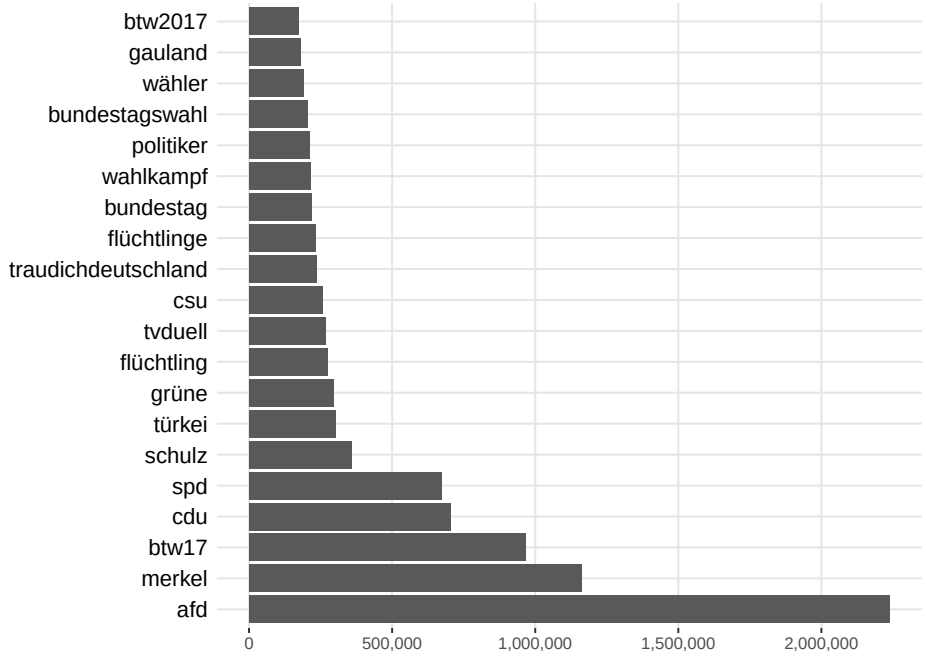


Figure 4: Top 20 mentioned selectors in our dataset.

larger parties (CDU, SPD, "groko" which stands for the governing "grand coalition" of the two) as well as their leading candidates Merkel and Schulz. We also find some generic terms like Bundestag or btw17, the foreign policy issue Turkey and the term refugees. Based on that figure, it seems that campaigning online resembles offline politics to a close extent – most attention is devoted to the leading parties and candidates (see also Yang & Kim, 2017), with the outlier being the AfD.

5.4 Activity and Resonance by Party

Public and academic observers have put a particular focus on the use of social media by political parties and candidates. Thus, our report relies on our exhaustive data to investigate how political actors themselves have used social media and how their activities were received by citizens.

In Table 2, we first list the number of candidates who were running in the campaign for the seven main parties, next the number of candidates for which we found at least one Facebook or Twitter account and finally for how many candidates data could be retrieved during our research period (i.e., a candidate had a public and active account). 84% of Bundestag candidates maintained a Facebook account, while 49% had a presence on Twitter. Moreover, for 710 candidates, we found a second Facebook account for which we also collected data. There is a considerable difference in the retrieval rates for Facebook (48%) and Twitter (67%). The lower retrieval rates on Facebook are due to the fact that many accounts were set up as user profiles or private accounts (instead of accounts declared as Facebook pages). In contrast, almost all Twitter accounts were public, but approximately one third of candidates with a Twitter presence never tweeted during the campaign.

words like "#noafd" or when these are part of mentioned screennames like @AfDBerlin.

Table 2: Social media adoption and number of retrieved accounts by party candidates

Party	Candidates	FB accounts	public & active FB	TW accounts	public & active TW
AfD	388	287	126	139	75
CDU	477	400	196	208	120
CSU	90	82	40	55	33
FDP	367	329	160	185	125
Grüne	360	313	114	212	174
Linke	355	290	123	163	109
SPD	479	408	252	261	181
Total	2,516	2,109	1011	1,222	817

We compared these results to the ones reported by Schmidt (2017). To be clear, the comparison is based on our final dataset that includes the candidate accounts we added from the Tagesspiegel and Open Knowledge Foundation databases. Schmidt and his team found a Facebook account (pages or profiles) for 1,849 candidates from the seven parties we focus on, compared to 2,109 in Table 2. He reports 1,096 candidates with a Twitter account, our final dataset contains 1,222 Twitter accounts.

Next, we turn to the resonance that political actors attract on social media. This gives indications on whether the considerable staff and resources needed to maintain a social media presence are worth the effort. For this, we use measurements of social media influence that are established in the literature (e.g., (Kovic et al., 2017; Nielsen & Vaccari, 2013; Yang & Kim, 2017). Note, however, that these metrics need to be interpreted with caution as they are particularly susceptible to automated activity by bots or astroturfing campaigns coordinated by humans (Keller, Schoch, Stier, & Yang, 2017). It also has to be kept in mind that deletions before the 8-day rolling time window of our data collection are not included here. Deletions are, however, less frequent for party posts than for user generated content (Bachl, 2017).

Figure 5 shows the top five accounts per party according to how many Facebook page likes they have accumulated. Accounts of leading politicians and the main party accounts dominate the list. When looking at the averages per party, we cannot find clear-cut differences. However, this count highly depends on the number of included accounts and how many candidates are on party lists. Overall, there are no apparent asymmetries between parties in the number of page likes they attracted.

In Figure 6, the same plot is displayed for the number of followers on Twitter, i.e., the people who regularly receive messages from political actors in their Twitter timelines. While Angela Merkel is the person with most page likes on Facebook, Martin Schulz leads the field in terms of Twitter followers. This is unsurprising, as Angela Merkel does not have a Twitter account and Schulz has gained international followers during his time as president of the European parliament. It is interesting to contrast the paltry follower numbers of leading and average AfD politicians to the prominent role of the AfD in the topics that were discussed on Twitter (see Figure 4). Their lower number of followers notwithstanding, the party still overshadowed Twitter debates through various escalations in campaign rhetoric. This indicates that even though the AfD dominates Facebook and Twitter discourses, the tone of these debates (support for the AfD vs. negative depictions) might be rather different.

In our final analysis, we move beyond the page level and investigate the actual engagement of audiences with contents produced by parties. This might be a better indicator for the "success" of online campaigns. As Figure 7 makes strikingly clear, the AfD was by far the most successful party on Facebook in terms of engagement. Even though AfD candidates and party branches only posted

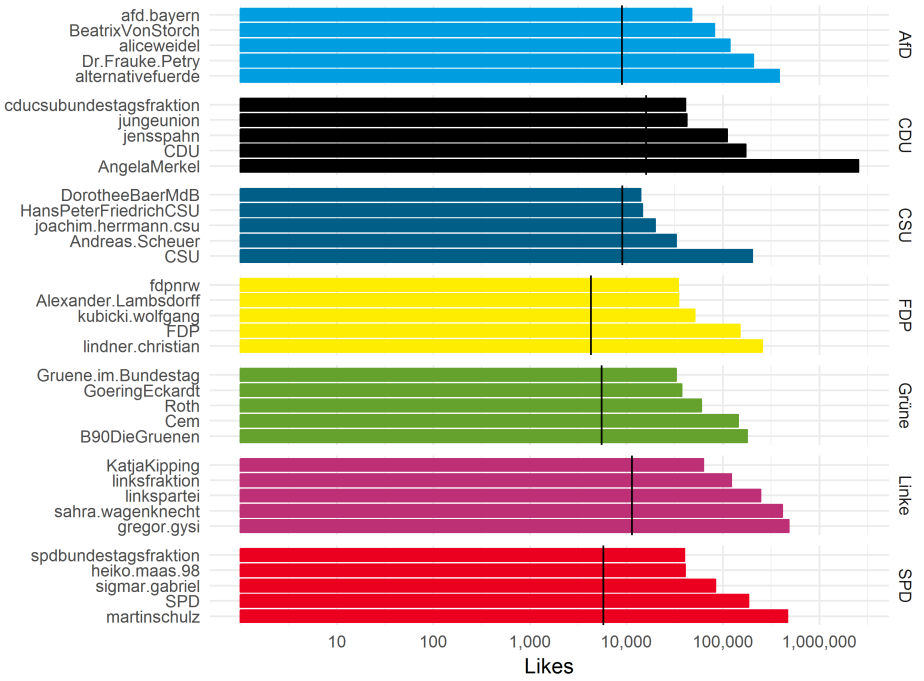


Figure 5: Top five Facebook accounts by page likes for each party on a logged scale. The black lines represent party averages.

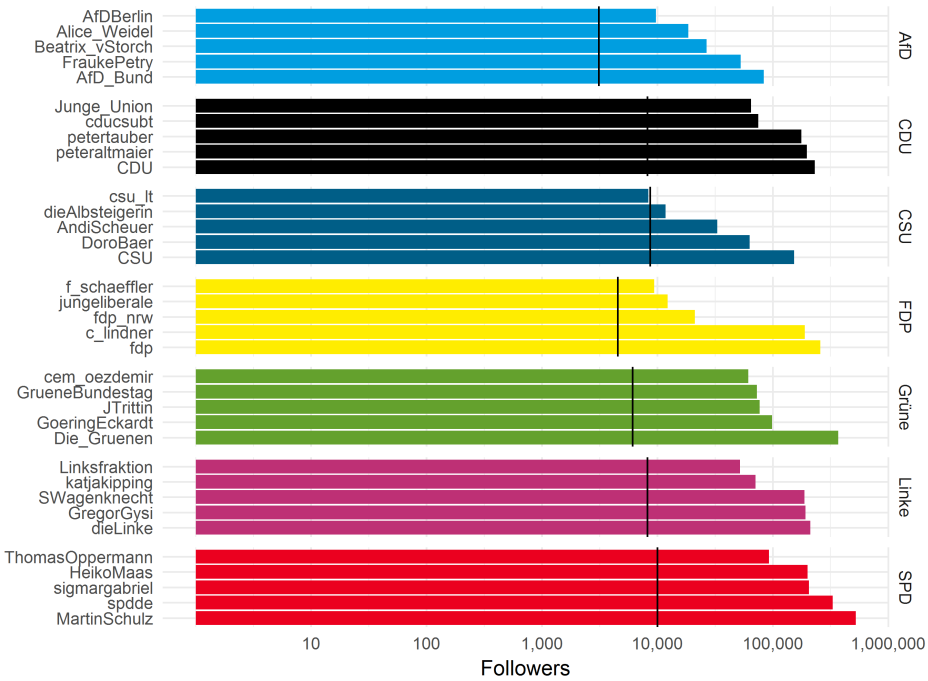


Figure 6: Top five Twitter accounts by the number of followers for each party on a logged scale. The black lines represent party averages.

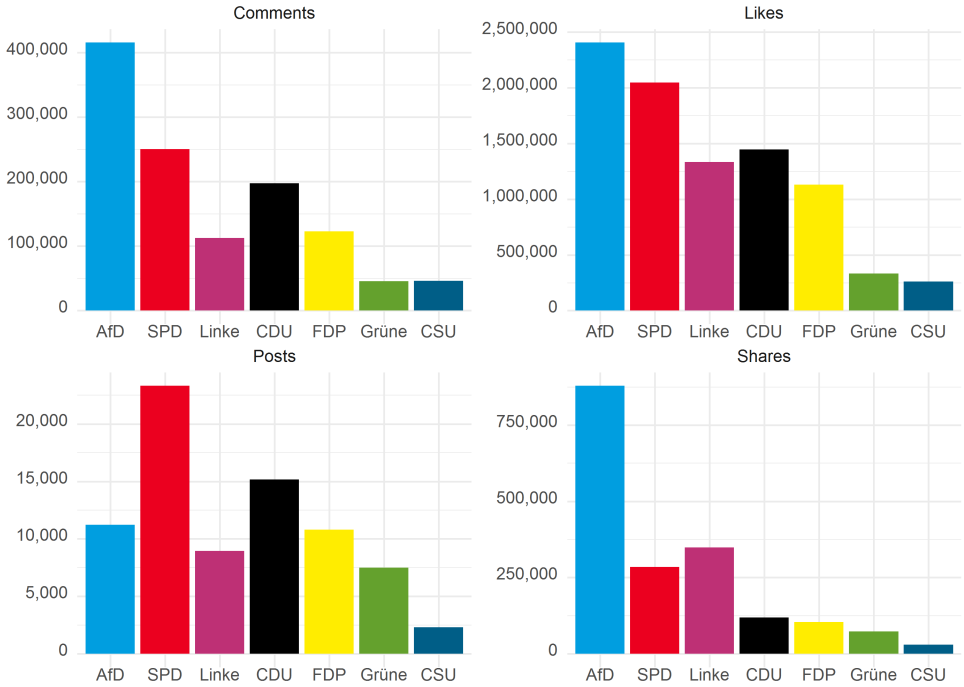


Figure 7: Aggregated Facebook posts and received engagement (comments, likes and shares) by party.

an average number of posts, they received higher numbers of likes, comments and in particular shares than other parties. Likes and shares can be regarded as the clearest indications for political support, while comments tend to be distributed more evenly since they are also used to criticize account holders.

Overall, these results suggest that the AfD was successful in generating the most (positive or negative) attention during the election campaign. This echoes results from previous studies (Neudert et al., 2017; Stier et al., 2017) and media reports on the AfD's extensive social media operations.[17] Further research should investigate how AfD accounts are intertwined with supporter groups and so-called "alternative media" pages.[18] There is mounting evidence that especially on Facebook, a right-wing media ecosystem has emerged that continues to influence German online discourses after the election campaign.

6 Data Sharing

The data we collected is proprietary data owned by Facebook and Twitter. Because of that and due to privacy restrictions, it is not possible to share the raw data with other researchers or the public. However, there are possibilities to reconstruct social media datasets, which is important in order to promote reproducibility in digital research (Kinder-Kurlanda, Weller, Zenk-Möltgen, Pfeffer,

[17] http://www.sueddeutsche.de/politik/gezielte-grenzverletzungen-so-aggressiv-macht-die-afd-wahlkampf-auf-facebook-1.3664785-2

[18] Account names of such online media can also be downloaded from Stier, Bleier, Bonart, et al. (2018).

Figure 8: Interface of the online monitoring tool.

& Morstatter, 2017). Twitter allows the publication of tweet IDs for academic purposes. Using these unique numeric identifiers, researchers can reconstruct our Twitter dataset. We publish our masterlist of accounts which allows for the reconstruction of the Facebook dataset – with certain limitations in ex post data availability (Bachl, 2017). The necessary data for the reconstruction of the Twitter and Facebook datasets can be found at Stier, Bleier, Bonart, et al. (2018).

While raw social media data cannot be shared, researchers can publish aggregated information from the metadata or text of social media posts. Thus, we provide an online access to our dataset that allows researchers to track the salience of topics of interest in customized searches. The resulting time series show how the salience of terms like "merkel" or "afd" develops over time on social media. In its current beta version, only posts by partisan actors, i.e., party branches and candidate accounts are included. In further versions, we will include data by regular users posting and commenting on political matters.

Figure 8 shows a screenshot of the monitoring platform with a time series from Facebook. We used keywords on the two most realistic coalition options "jamaika" and "grosse koalition" as search input in this example. The tool not only enables researchers to explore how social media reacts to unfolding political events, but also to download the data as relative frequencies over time. Moreover, the tool allows for a filtering of the data by party. The monitoring instrument can be accessed at mediamonitoring.gesis.org.

7 Conclusions

This project aimed to provide guidelines on how to collect social media data in a valid and reliable way, with a particular focus on the specifics of political communication. Conceptually, we first defined relevant target concepts prominent during an election campaign, namely candidates, organizations like party and media accounts and political topics. We then operationalized and collected data on these concepts in line with the possibilities and affordances of the two social media platforms Facebook and Twitter. Our temporal analysis showed that social media reacts considerably to campaign events like the TV debate or election night. In an analysis of party activities and their reception on social media, we found coherent evidence that the AfD was particularly successful

in exploiting the potential of social media. Besides accounts of leading figures from other parties, the AfD generated the highest engagement on Facebook and was also the most talked about party on Twitter. As the political role of social media continues to grow, researchers will have to continue focusing on these partisan asymmetries that can only be identified if the data collection goes beyond a tiny collection of party accounts or campaign hashtags. It is of utmost importance to analyze political communication across all arenas of election campaigning, also in the long tail of candidates.

Based on these conceptual and infrastructural foundations, we now maintain a robust data collection scheme that will continue to monitor social media in Germany. Beyond political communication, the infrastructure can be activated for collecting data on various events and social phenomena ranging from election campaigns in other countries to non-political matters. The datasets on the Bundestag campaign can be reconstructed using the data we uploaded. The ongoing streaming of data can be searched and accessed via our monitoring tool. We think that these efforts represent a considerable step towards enhancing replicability and data sharing in social media research.

We also acknowledge several limitations and the need for further infrastructural investments to fully realize the value of digital behavioral data in the social sciences. The observational data we collected is limited in that it does not contain important contextual and individual-level information such as socio-demographic characteristics, voting behavior or political preferences. The lack of such data limits the theoretical and empirical value of digital behavioral data. Thus, we see a lot of potential in the linking of social media datasets with additional data sources like traditional surveys (Jungherr et al., 2016; Stier, Bleier, Lietz, & Strohmaier, 2018) surveys of social media users (Vaccari et al., 2015), candidate surveys (Karlsen & Enjolras, 2016; Quinlan et al., 2017) or data from other sources such as real-time measurements during TV debates (Maier, Hampe, & Jahn, 2016). It is also clear that our previous data collections and the literature in general have mostly confined themselves to contentious political periods like election campaigns or offline protests. As a reaction to that, our data mining will continuously run throughout the current legislative session of the Bundestag. Having longitudinal data will allow social scientists to assess how every-day, less event-driven political processes manifest themselves on social media. Such data can serve as an analytical baseline and improve the understanding of whether established findings on online political communication still hold during less contentious political periods.

References

Aragón, P., Kappler, K. E., Kaltenbrunner, A., Laniado, D., & Volkovich, Y. (2013). Communication dynamics in Twitter during political campaigns: The case of the 2011 Spanish national election. *Policy & Internet*, *5*(2), 183–206. doi: 10.1002/1944-2866.POI327

Bachl, M. (2017). *An evaluation of retrospective Facebook content collection.* (2017 Annual Meeting of the DGPuK Methods Section)

Bruns, A., Burgess, J., et al. (2011). #ausvotes: How Twitter covered the 2010 Australian federal election. *Communication, Politics & Culture*, *44*(2), 37–56.

Caton, S., Hall, M., & Weinhardt, C. (2015). How do politicians use Facebook? An applied social observatory. *Big Data & Society*, *2*(2). doi: 10.1177/2053951715612822

DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLOS ONE*, *8*(11). doi: 10.1371/journal.pone.0079449

Driscoll, K., & Walker, S. (2014). Big data, big questions| working within a black box: Transparency in the collection and production of big Twitter data. *International Journal of Communication*,

8, 20.

Dubois, E., & Gaffney, D. (2014). The multiple facets of influence: Identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist*, *58*(10), 1260–1277. doi: 10.1177/0002764214527088

Freelon, D. (2017). Campaigns in control: Analyzing controlled interactivity and message discipline on Facebook. *Journal of Information Technology & Politics*, *14*(2), 168–181. doi: 10.1080/19331681.2017.1309309

Freelon, D., & Karpf, D. (2015). Of big birds and bayonets: Hybrid Twitter interactivity in the 2012 Presidential debates. *Information, Communication & Society*, *18*(4), 390–406. doi: 10.1080/1369118X.2014.952659

Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, *13*(1), 72–91. doi: 10.1080/19331681.2015.1132401

Jungherr, A., Schoen, H., & Jürgens, P. (2016). The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013. *Journal of Computer-Mediated Communication*, *21*(1), 50–68. doi: 10.1111/jcc4.12143

Jürgens, P., Jungherr, A., & Schoen, H. (2011). Small worlds with a difference: New gatekeepers and the filtering of political information on Twitter. In *Proceedings of the 3rd International Web Science Conference* (pp. 21–25). New York, NY: ACM.

Kaczmirek, L., Mayr, P., Vatrapu, R., Bleier, A., Blumenberg, M., Gummer, T., ... others (2014). Social media monitoring of the campaigns for the 2013 German Bundestag elections on Facebook and Twitter. *GESIS – Working Papers*, *31*. Retrieved 10 July 2017, from https://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_arbeitsberichte/WorkingPapers_2014-31.pdf

Karlsen, R., & Enjolras, B. (2016). Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with Twitter data. *The International Journal of Press/Politics*, *21*(3), 338–357. doi: 10.1177/1940161216645335

Keller, F., Schoch, D., Stier, S., & Yang, J. (2017). How to manipulate social media: Analyzing political astroturfing using ground truth data from South Korea. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* (pp. 564–567). Menlo Park, CA: The AAAI Press.

Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, *4*(2). doi: 10.1177/2053951717736336

King, G., Lam, P., & Roberts, M. E. (2017). Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science*, *61*(4), 971–988. doi: 10.1111/ajps.12291

Kovic, M., Rauchfleisch, A., Metag, J., Caspar, C., & Szenogrady, J. (2017). Brute force effects of mass media presence and social media activity on electoral outcome. *Journal of Information Technology & Politics*, *14*(4), 348–371. doi: 10.1080/19331681.2017.1374228

Kratzke, N. (2017). The #btw17 Twitter dataset – recorded tweets of the federal election campaigns of 2017 for the 19th German Bundestag. *Data*, *2*(4). doi: 10.3390/data2040034

Lev-On, A., & Haleva-Amir, S. (2016). Normalizing or equalizing? Characterizing Facebook campaigning. *New Media & Society*. doi: 10.1177/1461444816669160

Lietz, H., Wagner, C., Bleier, A., & Strohmaier, M. (2014). When politicians talk: Assessing online conversational practices of political parties on Twitter. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (pp. 285–294). Palo Alto, CA: AAAI Press.

Lin, Y.-R., Keegan, B., Margolin, D., & Lazer, D. (2014). Rising tides or rising stars? Dynamics of shared attention on Twitter during media events. *PLOS ONE*, *9*(5). doi: 10.1371/journal.pone.0094093

Maier, J., Hampe, J. F., & Jahn, N. (2016). Breaking out of the lab: Measuring real-time responses to

televised political content in real-world settings. *Public Opinion Quarterly*, *80*(2), 542–553. doi: 10.1093/poq/nfw010

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media* (pp. 400–408). Palo Alto, CA: AAAI Press.

Neudert, L.-M., Kollanyi, B., & Howard, P. N. (2017). *Junk news and bots during the German parliamentary election: What are German voters sharing over Twitter?* COMPROP DATA MEMO 2017.7, Oxford University. Retrieved 06 January 2018, from http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/09/ComProp_GermanElections_Sep2017v5.pdf

Nielsen, R. K., & Vaccari, C. (2013). Do people "like" politicians on Facebook? Not really. Large-scale direct candidate-to-voter online communication as an outlier phenomenon. *International Journal of Communication*, *7*.

Quinlan, S., Gummer, T., Roßmann, J., & Wolf, C. (2017). Show me the money and the party! Variation in Facebook and Twitter adoption by politicians. *Information, Communication & Society*. doi: 10.1080/1369118X.2017.1301521

Russell Neuman, W., Guggenheim, L., Mo Jang, S., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, *64*(2), 193–214. doi: 10.1111/jcom.12088

Schmidt, J. (2017). Twitter-Nutzung von Kandidierenden der Bundestagswahl 2017. *Media Perspektiven*, *12/2017*, 616–629.

Stier, S., Bleier, A., Bonart, M., Mörsheim, F., Bohlouli, M., Nizhegorodov, M., ... Staab, S. (2018). *Social media monitoring for the German federal election 2017.* GESIS Data Catalogue DBK. Retrieved 27 February 2018, from http://dx.doi.org/10.4232/1.12992

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences and the mediation of political communication on Facebook and Twitter. *Political Communication*, *35*(1), 50–74. doi: 10.1177/1461444817709282

Stier, S., Posch, L., Bleier, A., & Strohmaier, M. (2017). When populists become popular: Comparing Facebook use by the right-wing movement Pegida and German political parties. *Information, Communication & Society*, *20*(9), 1365–1388. doi: 10.1080/1369118X.2017.1328519

Strohmaier, M., & Wagner, C. (2014). Computational social science for the world wide web. *IEEE Intelligent Systems*, *29*(5), 84–88. doi: 10.1109/MIS.2014.80

Trilling, D. (2015). Two different debates? Investigating the relationship between a political debate on TV and simultaneous comments on Twitter. *Social Science Computer Review*, *33*(3), 259–276. doi: 10.1177/0894439314537886

Tromble, R., Storz, A., & Stockmann, D. (2017). We don't know what we don't know: When and how the use of Twitter's public APIs biases scientific inference. *SSRN*. Retrieved 30 December 2017, from https://ssrn.com/abstract=3079927

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media* (pp. 178–185). Palo Alto, CA: AAAI Press.

Vaccari, C., Chadwick, A., & O'Loughlin, B. (2015). Dual screening the political: Media events, social media, and citizen engagement. *Journal of Communication*, *65*(6), 1041–1061. doi: 10.1111/jcom.12187

Vergeer, M., Hermans, L., & Sams, S. (2013). Online social networks and micro-blogging in political campaigning: The exploration of a new campaign tool and a new campaign style. *Party Politics*, *19*(3), 477–501. doi: 10.1177/1354068811407580

Williams, C. B., & Gulati, G. J. (2013). Social networks in political campaigns: Facebook and the congressional elections of 2006 and 2008. *New Media & Society*, *15*(1), 52–71. doi: 10.1177/1461444812457332

Yang, J., & Kim, Y. M. (2017). Equalization or normalization? Voter–candidate engagement on

Twitter in the 2010 U.S. midterm elections. *Journal of Information Technology & Politics*, *14*(3), 232–247. doi: 10.1080/19331681.2017.1338174

A Appendix

Table A1: Selectors: **Polity**

bundestag	bundesministerin	direktmandat
länderfinanzausgleich	bundesminister	wahlrecht
bundeskanzler	landtag	landtagswahl
kanzler	bürgerentscheid	verfassungsgericht
bundeskanzlerin	wahlsystem	bundesverfassungsgericht
direktedemokratie	bundesrat	bverfg
direkte demokratie	bundespräsident	bundesregierung
überhangmandat	wähler	verfassungsrichter
bundestagswahl		

Table A2: Selectors: **Policy**

finanzpolitik	zensur	energiewende
steuerpolitik	shadowbans	klimawandel
staatshaushalt	shadow bans	kulturpolitik
wirtschaftslage	atomkraft	schulpolitik
russland	atomenergie	einkommensungleichheit
türkei	antiatom	steuergeschenk
syrien	endlagerung	wirtschaftspolitik
arbeitslosigkeit	gorleben	integrationspolitik
steuerreform	atomausstieg	rentenpolitik
vorratsdatenspeicherung	überwachung	betreuungsgeld
bankenkrise	datenschutz	frauenquote
finanzkrise	überwachung	einkommensgerechtigkeit
ezb	islamist	fachkräftemangel
bankenaufsicht	lohnpolitik	gesundheitspolitik
finanzmarktsteuer	ehefueralle	bildungspolitik
agenda 2010	ehefüralle	kinderarmut
herdprämie	staatsdefizit	arbeitsmarktpolitik
einkommensschere	ökosteuer	altersarmut
jugendarbeitslosigkeit	hochschulpolitik	solidaritätszuschlag
forschungspolitik	hartzIV	auslandseinsatz
verkehrspolitik	hartz4	elterngeld
familienpolitik	hartz 4	freihandel
bundeswehr	sozialpolitik	flüchtling
umweltpolitik	soziale gerechtigkeit	fluechtling
rentenreform	mindestlohn	flüchtlinge
netzpolitik	wirtschaftskrise	fluechtlinge
Netzwerkdurchsetzungsgesetz	klimaschutz	migranten
netzdg	energiepolitik	

Table A3: Selectors: **Politics / election campaign**

wahl17	wahljahr	infratest
btw2017	umfrage	wahlbeteiligung
btw17	pegida	populismus
wahlen2017	forschungsgruppe wahlen	populistisch
wahl2017	lügenpresse	politikverdrossenheit
kanzlerduell	luegenpresse	wahlversprechen
tvduell	forsa umfrage	politiker
elefantenrunde	politbarometer	wahlwerbung
parteiprogramm	emnid	meinungsfreiheit
schulzzug	allensbach	wahlkampf
parteitag		

Table A4: Selectors: **Parties & leading politicians**

grüne	karrenbauer	steinmeier
bündnis90	altmaier	thorsten albig
buendnis90	bouffier	woidke
gruene	groehe	#fdp
goering-eckardt	rainer haseloff	lindner
göring-eckardt	stanislaw tillich	npd
goeringeckardt	von der leyen	piratenpartei
göringeckardt	gröhe	koalition
kretschmann	linkspartei	groko
oezdemir	bartsch	grosse koalition
özdemir	wagenknecht	große koalition
hofreiter	ramelow	jamaikakoalition
csu	spd	ampelkoalition
seehofer	schulz	schwampel
dobrindt	hannelore kraft	schwarzgrün
christianschmidt	michaelmüller	schwarzgelb
afd	heiko maas	rotgrün
frauke petry	carsten sieling	fedidwgugl
gauland	schwesig	traudichdeutschland
weidel	sigmar gabriel	holdirdeinlandzurück
cdu	barbara hendricks	denkenwirneu
merkel	sellering	lustauflinks
schäuble	olaf scholz	darumgrün
schaeuble	malu dreyer	darumgruen
maiziere	nahles	zeitfürmartin
johanna wanka	stephan weil	zeitfuermartin