From *Alt-Right* to *Alt-Rechts*: Twitter Analysis of the 2017 German Federal Election

Fred Morstatter
University of Southern California
Information Sciences Institute
Marina Del Rey, California
fredmors@isi.edu

Yunqiu Shao
Tsinghua University
Department of Computer Science & Technology
Beijing, China
shaoyunqiu14@gmail.com

Aram Galstyan
University of Southern California
Information Sciences Institute
Marina Del Rey, California
galstyan@isi.edu

Shanika Karunasekera
The University of Melbourne
Melbourne, Australia
karus@unimelb.edu.au

ABSTRACT

In the 2017 German Federal elections, the "Alternative for Deutschland", or AfD, party was able to take control of many seats in German parliament. Their success was credited, in part, to their large online presence. Like other "alt-right" organizations worldwide, this party is tech savvy, generating a large social media footprint, especially on Twitter, which provides an ample opportunity to understand their online behavior. In this work we present an analysis of Twitter data related to the aforementioned election. We show how users self-organize into communities, and identify the themes that define those communities. Next we analyze the content generated by those communities, and the extent to which these communities interact. Despite these elections being held in Germany, we note a substantial impact from the English-speaking Twittersphere. Specifically, we note that many of these accounts appear to be from the American alt-right movement, and support the German alt-right movement.

CCS CONCEPTS

• **Networks** → **Social media networks**; **Online social networks**;
• **Social and professional topics** → *Socio-technical systems*;

KEYWORDS

Social Networks, Online Campaigns, Bots

ACM Reference Format:
Fred Morstatter, Yunqiu Shao, Aram Galstyan, and Shanika Karunasekera. 2018. From *Alt-Right* to *Alt-Rechts*: Twitter Analysis of the 2017 German Federal Election. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3184558.3188733

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3188733

1 INTRODUCTION

Politics is a widely-discussed topic on social media platforms. On Twitter, the US election alone was the second-most tweeted event of 2016.[1] In fact, social media is so important to the outcome of elections that many parties spend ample amounts of money on their online campaigns. While all parties focus on social media, some see more impact than others. One of the ideologies to benefit most from social media is the "alt-right," which is "a set of ideologies, groups, and individuals" who oppose multiculturalism and immigration.[2] Recent research has shown that they are nationalists and populists [9]. The alt-right finds its home online because the beliefs they espouse are not popular in mainstream media. The pseudo-anonymity offered by many social media platforms allows them to promote their views through the spreading of memes, and in doing so to recruit others to their cause. The most visible case of this was the 2016 US Presidential Election [10], however this has also been noted in other alt-right movements, such as the 2017 French Presidential Election [5].

In this work we study online discourse in the German Twittersphere leading up to the 2017 federal election. This election is of interest for a number of reasons. First, Twitter has not had the level of penetration that is has had in other countries. As evidence, the current Chancellor, Angela Merkel, does not have a Twitter account.[3] Second, this election was highly covered in the media because of the German party "Alternative for Germany" (AfD) who wished to seize larger control of German parliament. What brings these disparate points together is that AfD is "alt-right." As with other alt-right movements, such as the alt-right movement in the US [10], it largely started online and gained momentum through online discourse.[4] It is likely that the alt-right segment of the election discussion could be disproportionately large, and tilt Twitter in its favor.

[1] http://www.independent.co.uk/news/twitter-most-tweeted-moments-2016-donald-trump-brexit-black-lives-matter-rio-a7466236.html
[2] https://www.splcenter.org/fighting-hate/extremist-files/ideology/alt-right
[3] https://www.bloomberg.com/news/articles/2017-06-26/differences-aside-merkel-can-t-help-peeking-at-trump-s-tweets
[4] https://www.theguardian.com/commentisfree/2017/sep/26/germany-far-right-election-afd-trolls

In this work, we assess several aspects of political discourse on social media as it pertains to the 2017 German elections. We address the following questions:

(1) How do users organize into communities? How does information flow between them?
(2) What topics emerge from the text? How do these topics relate to the communities and how are they used throughout time?
(3) What role, if any, did bots play in the election?

These observations contribute to a larger understanding of how social media dynamics play out during elections.

2 2017 GERMAN ELECTION

The 2017 German Federal Election was held on September 24, 2017. This election received widespread attention because of the emergence of the AfD party. Previously, this party held no seats in German parliament. After the election, they held 94 seats. Their success was due, in part, to their large presence on social media.[5] In this work, we investigate how that presence manifested on social media, namely Twitter. To do this, we first collected data from Twitter pertaining to the election. Next, we analyze the communities and topics that presented themselves in the days leading up to the election. We continue to measure the footprint of bots.

2.1 Obtaining Election-Related Twitter Data

We identified users who posted tweets related to the election from June 27 to August 7, 2017. We observed those users who were tweeting hashtags pertaining to the German election: #btw17, #afd, #cdu, #spd. The latter three hashtags represent major political parties while the first was the main hashtag used to discuss the election. This processed yielded 6,139 users. We crawled these users from August 10 - September 24, 2017. We collected their data using Twitter's Filter API[6] from two perspectives. First, by monitoring the messages they posted through the "follow" parameter. Second, we wanted to obtain all of the retweets of their tweets. This was accomplished by using each user's screen name as a keyword to the "track" endpoint. All tweets collected through the second mechanism were further verified to ensure they are actually retweets and not simply keyword mentions. At the end of our data collection process, the dataset consisted of 34,383,201 tweets from a total of 133,098 users. The data will be published on acceptance of this work in accordance with Twitter's terms of service.

3 ORGANIZATION OF COMMUNITIES

In this section we investigate the properties of these users via the data we collected. Who are they? How are they connected? What communities exist in this network and what do they represent? Our analysis will help to shed light on who these users are.

3.1 Discovering Communities

To extract communities from the network, we construct a network based on the re-tweets and mentions of all the users we tracked.

Figure 1: User community assignments. Green: Community 4, Purple: Community 3, Blue: Community 1, and Orange: Community 8.

We chose these communication channels as they both represent a reaction from one user to another. Using the Louvain algorithm [15], we decomposed this network into several different communities.

The result of the community detection algorithm is shown in Figure 1. While the algorithm discovered 18 communities, only four of those are major communities containing more than 400 users. We focus on these four meaningful communities in the figure. Hereafter, these communities will be referred to by their modularity class. We obtain four main communities: the green one on the left side (modularity class 4), the purple one on the upper-right (modularity class 3), the blue one on the bottom (modularity class 1) and the orange one in the middle (modularity class 8). At first, we looked at some basic characteristics in quantity of these communities in Table 1. We consider users who tweet more than 300 times a day for at least 5 days as extremely active users. The count of extremely active users by community is shown in the last column of Table 1. Community 4 has many more active users than the other three communities considering the average number of tweets per day and the proportion of extremely active users. Community 1, though it contains the fewest users, is quite an active community among the four. We considered tweets authored in German as well as English as German is the official language of Germany but we were also interested in the opinions of others living in and outside the country [14]. The number (and percentage) of English tweets and German tweets of each community (from 10th August to 24th September) is shown in Table 2. It clearly indicates that Community 4 is an English-speaking community while Community 1, Community 3 and Community 8, are the three German-speaking communities.

4 COMMUNITY CROSS POLLINATION

We have established that there are four major communities in the election. We want to know the extent to which information flows between these communities. Is community 4, our English-speaking,

[5] https://www.theatlantic.com/international/archive/2017/09/afd-harris-merkel-germany-elections/541506/
[6] https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter

Table 1: Number of users in each community alongside their average number of tweets per day, number (percentage) of extremely active users in each community.

Community	Users	Tweets/Day	Extremely Active Users
C1	479	18139	18 (3.76%)
C3	1182	43757	35 (2.96%)
C4	935	84942	133 (14.22%)
C8	613	22660	16 (2.61%)

Table 2: The number (proportion) of English tweets vs. number (proportion) of German tweets in each community.

Community	German	English
C1	700,285 (92.1%)	60,346 (7.9%)
C3	1,175,821 (72.9%)	437,132 (27.1%)
C4	100,152 (3.5%)	2,776,117 (96.5%)
C8	693,375 (76.7%)	210,906 (23.3%)

Table 3: The number of retweets between communities. The row indicates the retweeting community, and the column indicates the retweeted community.

	1	3	4	8
1	176,055 (55%)	40,085 (13%)	12,948 (4%)	90,194 (28%)
3	24,613 (9%)	182,935 (70%)	8,528 (3%)	45,652 (17%)
4	13,492 (3%)	9,412 (2%)	429,037 (92%)	16,050 (3%)
8	53,416 (20%)	39,115 (15%)	12,288 (5%)	163,329 (61%)

alt-right community an island, or does it actually have an effect on the dialog in the German election? To answer this question, we measure the extent to which users in each community retweet members of their own community, and to the extent that they retweet members of other communities. The results are shown in Table 3. As expected, the retweets tend to happen within communities.

Nevertheless, two conclusions can be drawn from this table. First, there is a strong relationship between Communities 1 and 8. C1 is the alt-right German-speaking community, and C8 mostly pertains to general election discussion so it is unsurprising that these two interact. Second, the English-speaking alt-right community does not have much interplay with the other communities in the graph. These users neither retweet, nor are they retweeted by, members of other communities. It is unlikely that this is due to ideological differences, as Community 1 shares many of the beliefs. Instead, it is likely that this is a byproduct of language barriers between the two groups. This pattern also emerged in the 2017 French presidential election, where communication barriers are cited as having an effect on the alt-right's ability to penetrate the discussion [19].

5 FROM COMMUNITIES TO TOPICS

While understanding the communities active in the election is important, it is also important to analyze what they say. In this section, we focus on the topics of focus for these communities. The hashtags used in a tweet are indicators of the topic of the tweet. However, even when reducing the tweet to a set of hashtags, the volume of both tweets and hashtags makes understanding the text intractable. Instead, we investigate the clusters of keywords in the form of topics.

Total Correlation Explanation (CorEx) [21, 23] is an information-theoretic unsupervised algorithm to learn representations of data that "explain" dependence in the data. It constructs a hierarchy of latent factors that progressively explain more dependencies in the observations as measured by multivariate information (total correlation). We used an implementation of CorEx, corex_topic[7], which is developed for topic modeling using binary bags of words and can easily handle thousands of variables, which is consistent with the vocabulary of our dataset. We use this for topic modeling on hashtags. To leverage CorEx, we consider each user as a 'document', and each hashtag as a 'term'. We extracted hashtags from tweets from the entirety of the dataset. We filtered out the hashtags that appeared less than 30 times every day to avoid noise caused by rare hashtags. As a result, we extracted 3,075 hashtags of 3,209 users in all 4 communities. Using this data we constructed a bipartite graph with users and hashtags as nodes and an edge representing the user posting the hastag in at least one of his/her tweets. We used this graph as input of CorEx for topic modeling and obtained a hierarchical structure of 2 layers of all the hashtags. For the first layer, there were 50 topics, and for the second layer, there were 7 clusters. Figure 2 shows the underlying structure.

To summarize the topics shown in Figure 2, the topics fall into the following categories:

- Topics directly related to German election: parties, candidates, election-related propaganda activities, etc.
- Social problems: refugees, migration, islam, racism, terrorism, welfare, democracy, fake news, safety, etc.
- Elections in other countries: US election, Brexit
- Other countries and organizations: USA, Turkey, North Korea, Italy, France, EU, etc.
- Social media: Twitter, Facebook, broadcast, press, etc.
- Other non-political topics: football, traveling, hurricane,etc.

In the higher level topics, which are marked by red text in Figure 2, there are seven clusters (0-6). Cluster 0 contains topics about elections both in and outside Germany as well as topics about social problems, while Cluster 1 and Cluster 4 are topics mostly focused on social problems in Germany. Cluster 2 topics are mostly about the USA and its social problems. Cluster 3 contains mostly topics about European countries and their social issues. Cluster 6 topics are directly related to German election. Cluster 5 contains mainly non-political topics such as animals, hurricane, etc.

There are numbers of hashtags which are acronyms. For example, #cdu, #spd, #fdp, #afd, #grüne are all active Parties in Germany. #btw17 is the acronym for "DeutscheWahl 2017" (German election 2017). For more information, Angela Merkel (represented as #merkel) and Martin Schulz (#schulz) are two candidates for Chancellor, and Merkel belongs to CDU while Schulz belongs to SPD. The Party AfD (#afd) is a growing right-wing Party, which attracted a large amount of attention from social media users. #traudichdeutschland represents the slogan for it and Alexander Gauland (#gauland) is one of the founders of it.

[7] https://github.com/gregversteeg/corex_topic

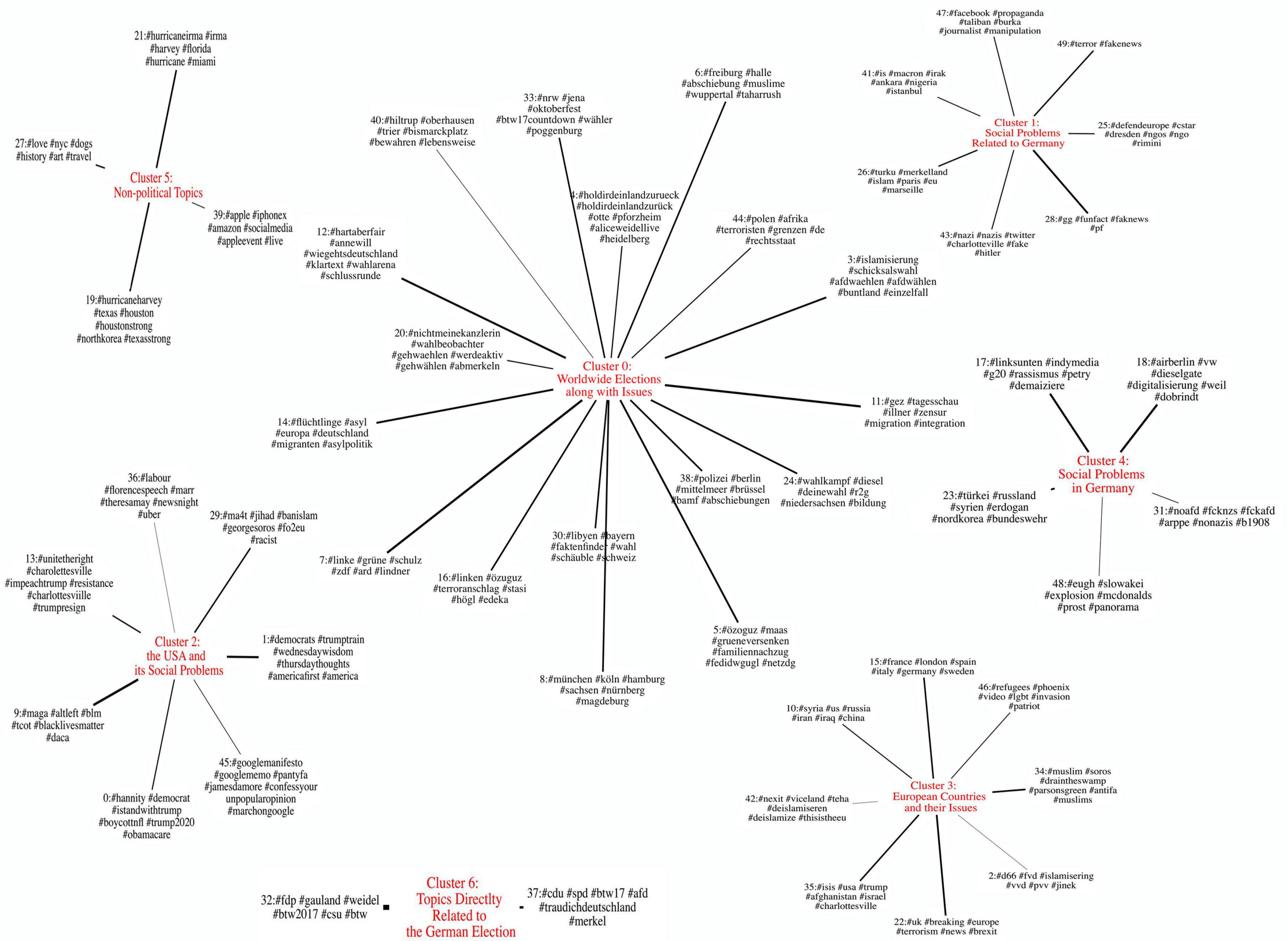
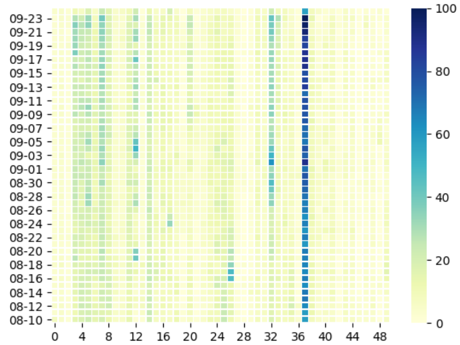


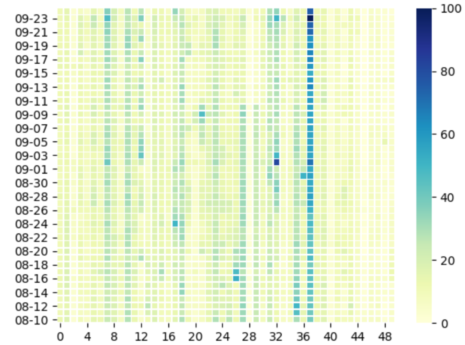
Figure 2: The hierarchical structure of topics from CorEx. The first layer has 50 topics, and the figure shows top 6 hashtags in each topic. The red text indicates the second layer. The 50 topics are divided into 7 clusters in higher level.

How the communities differ in topics: In essence, we have two clusterings. The first is the communities, the clustering of users based upon their position in the social network. The second is the topics, a clustering of text based upon redundancies introduced by the users. In this section we study the interaction of these clusterings. To measure this, we define a metric to gauge the communities focus, or attention, to a particular topic. We define the attention score of topic, $attent$, to measure how much attention each topic is given by the communities. Initially, $attent_i = 0\ for\ i = 0, 1, \ldots, 49$. $attent_j = attent_j + 1$ means the hashtag in Topic j is used for one time. We performed this analysis based on the daily tweets of users and generated the attention score heat maps for 4 communities as shown in Figure 3. The x-axis is the topic, the y-axis is the date. The color represents the roots of attention scores. The attention scores for topics (e.g. Topic 37, Topic 32) directly related to the German election showed an obvious increase for the three German-speaking communities, especially for Topic 37, which contains the parties the two candidates belong to (#cdu, #spd) and AfD-related tags (e.g. #afd, #traudichdeutschland). In Community 4 (the English-speaking community), Topic 37 also received a growing attention score with time, however, compared with other topics, the score was not very high and the increasing tendency was not as obvious as that in the other three communities. Community 4 seemed concerned with different topics compared to the other groups, e.g. Topic 9, Topic 0, Topic 22, Topic 34, Topic 35 etc. For Community 4, there did exist some sharp peaks in scores for particular topics in specific days e.g. Topic 13, 26, which are mostly about hot-button issues all over the world and have some relation to terrorism and conflict. This suggested that large numbers of users in this community were active in talking about social problems in the world and might have a right-wing tendency. The German Election was just part of their attention in this period. The 5 most frequent hashtags in the topics we have identified are shown in Table 4.

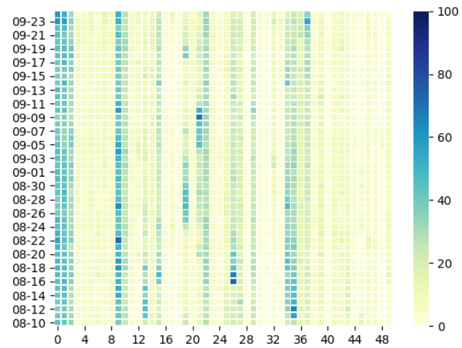
To be more specific, we looked at the hashtags related to the three parties: #afd, #cdu and #spd. Although #afd is the most frequently used for all communities, Community 1 had extremely



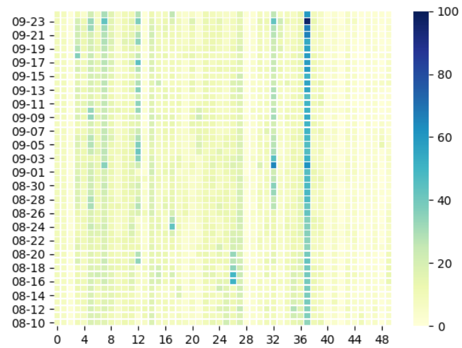
(a) Community 1



(b) Community 3



(c) Community 4



(d) Community 8

Figure 3: Daily attention scores of each topic for the four communities. The x-axis is the topic, the y-axis is time and the color represents the roots of the attention score. For consistency, the highest number is limited to 100 in these figures while some attention scores of Topic 37 in Community 1 and Community 3 reach about 120 actually.

high propensity to use #afd. On average, Community 1 used #afd about 2,500 times a day, much higher than other communities even though it contained the fewest users. For Community 1, #afd appeared about 10 times more than the other two hashtags, while for the other 3 communities it was about 2 or 3 times. The appearance of #traudichdeutschland(the slogan for AfD) and #noafd also drew our interests. Community 1 used #traudichdeutschland over 1000 times a day. However, we could seldom find #noafd in Community 1's tweets. On the contrary, for Community 3, these two hashtags appeared with nearly the same frequency, and for Community 8, they both seldom appeared. It was quite abnormal indicating that Community 1 could be alt-right and have a great propensity to support AfD. There could also exist some organizations which tried to manipulate Twitter for specific political purposes.

Table 4: The five most frequent hashtags in the topics mentioned in this section. We use (e.g.) T0 as an abbreviation for Topic 0.

Topic No.	Top 5 tags
T0	#livehealingly, #tsha, #boycottnfl, #istandwithtrump, #dems
T3	#afdwaehlen, #schicksalswahl, #islamisierung, #einzelfall, #afdwÃďhlen
T9	#maga, #daca, #tcot, #obama, #pjnet
T13	#unitetheright, #altright, #impeachtrump, #lookatme, #charolettesville
T22	#breaking, #brexit, #uk, #europe, #rt
T26	#barcelona, #islam, #eu, #merkelland, #paris
T27	#history, #nyc, #love, #quote, #travel
T32	#btw2017, #tvduell, #fdp, #gauland, #weidel
T34	#antifa, #muslim, #draintheswamp, #koran, #soros
T35	#trump, #charlottesville, #isis, #usa, #daesh
T37	#afd, #btw17, #merkel, #traudichdeutschland, #spd

5.1 Community Topic Variation over Time

Next, we want to see how the attention changed with time. We used the daily binary User-Topic vector. $\boldsymbol{V_{icu}}$ is the vector for user u in community c on the i^{th} day, if he uses a hashtag in Topic j that day, then $v_{icuj} = 1$, otherwise $v_{icuj} = 0$. Then we computed the cosine similarity score sim_{icu} of $\boldsymbol{V_{icu}}$ and $\boldsymbol{V_{(i+1)cu}}$, that is $sim_{icu} = \dfrac{\boldsymbol{V_{icu}} \cdot \boldsymbol{V_{(i+1)cu}}}{||\boldsymbol{V_{icu}}||_2 ||\boldsymbol{V_{(i+1)cu}}||_2}$. We used the average of similarity score of all users in a community as the similarity score for this community Sim_{ic}. We also randomly select users from the whole set as randomly-generated communities to compute the similarity score. We repeat this random process 1000 times and use the average similarity score as the baseline to test the significance of our findings. Figure 4 shows how the similarity score of the 4 communities as well as the random-generated community changed daily and the second column of Table 5 shows the average similarity score SIM_c for this period. From the Figure 4, the similarity score lines for Community 1, Community 3 and Community 8 as well as the random-generated community all show a trend of growth as a whole, which indicates that users in these communities tended to increasingly focus on some specific topics with time. Combined with the heatmaps of Community 1, Community 3 and Community 8 in Figure 3, all of these communities show a growing attention on Topic 37, the topic directly related to the German election. It also explains that as the election get closer, the communities tend to become more and more similar. It is consistent with the assumption above that people in Germany were becoming more and more engrossed in election-related talking points. But the similarity of Community 4 goes down after 15 and doesn't show an increasing tendency like other communities. The attention heatmap for Community 4 Figure 3(c) does not show an obvious or extreme increase in Topic 37 either. It indicates that this community is not always focused on specific topics like the 3 Germany-speaking communities with the German election going on. The similarity score line of Community 4 is always above others because of the diversity in topics they paid attention to, which also can be confirmed by the Heatmaps in Figure 3. It suggests that Community 4 was an issue-driven community, which means they showed interest in diverse hot topics (most of them are about social issues) all over the world. When there is some startling news, they will focus on the related topics on specific days, which explains the great fluctuation this community has.

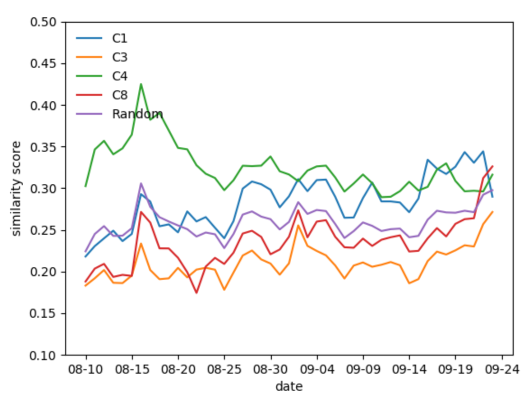


Figure 4: The similarity scores of the communities.

Another observation about Figure 4 is that from August 10th to 15th, there was an increase in activity for all communities that reached a peak on the 15th. We looked at the top hashtag sets of each community and we found #charlottesville. There was an attack in Charlottesville in the USA, which was a white supremacist event.[8] This event echoed on social media for several days. While this was going on, racism was the topic of concern in Europe, including Germany. It also involved other similar issues in Europe, i.e. the Defend Europe organization (#defendeurope). Racism and immigration are serious social issues in Germany. During the election, it was an inevitable topic for all parties and politicians. Thus, it caused a hot discussion all over the world along with other topics related to racism. It may explain why people in all four communities tend to continuously stay consistent.

[8]http://time.com/charlottesville-white-nationalist-rally-clashes/

In this part, we define users whose average similarity score in this period is > 0.7 as topic-consistent ones. The last column of Table 5 shows the number of topic-consistent users in each community. Looking at Table 5, Community 3 has the lowest average similarity score as well as the lowest percentage of topic-consistent users. Users in this community tend to look at diverse topics and change their topics irregularly day by day, but with the day of German election approaching, they became more focused on the topics related to the German election. The similarity scores (both average and daily) of Community 1 and Community 4 are significantly higher than others. Moreover, they have a much higher proportion of topic-consistent users, which indicates there may exist some organized users in them who are more engrossed in specific topics. The similarity score of Community 1 is always the highest among the German-speaking communities. Unlike Community 4, topics of Community 1 show little diversity. Combined with the significantly highest proportion of topic-consistent users in Community 1, it indicated that users in this community were focused on the same topics. It is likely that there did exist some political organizations in it that insisted on using Twitter for election-related purposes. The shape of the line of Community 8 in Figure 4 is similar to Community 3 and its average similarity score as well as the percentage of topic-consistent users is just a little bit higher than that of Community 3. The difference may lie in that they were usually more concerned about topics related to society and politics in Germany than users in Community 3.

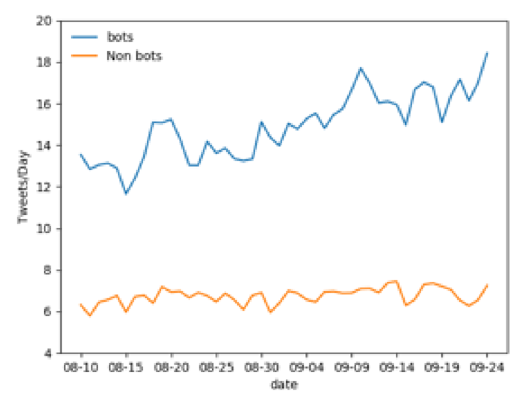


Figure 5: Bot activity over time. Bots create more noise per capita, and as the election nears they increase their activity.

In summary, Community 1 was the most suspicious. There was some evidence indicating that this community might contain social bots that manipulate social media and some latent campaigns, and might have an obvious tendency for the AfD Party. Community 3

Table 5: The average similarity score and the number of topic-consistent users in each community during this period.

Community	Avg. similarity score	> 0.7
C1	0.2831	15 (9.39%)
C3	0.2100	10(0.85%)
C4	0.3246	64(6.84%)
C8	0.2349	10(1.63%)
random-generated	0.2592	—

and Community 8 seemed like normal German-speaking communities. The difference between these two, as indicated by Figure 3 may lie in that users in Community 8 were more concerned about topics related to German society and politics, while users in Community 3 contained more people that took interest in diverse topics about all aspects of everyday life. Different from the above three communities, Community 4 was an English-speaking community that in addition to the German election, paid great attention to social problems all over the world. This community was active, and issue-driven.

6 THE IMPACT OF SOCIAL BOTS

Social media platforms are flooded with automated accounts, called "bots," many of which attempt to skew the opinion of real users by inundating them with information that matches the bots' goals [7]. This is especially prevalent in the context of politics where bots often try to skew the perception of a candidate, either in terms of inflating their perceived popularity (e.g., retweeting or following the candidate automatically), or posting content in favor of the candidate [8, 20]. Due to the suspicious nature of some of the users in the communities we detected, we hypothesize that there may be social bots operating in our dataset. In this section, we will estimate the impact of bots.

First we needed to obtain ground truth about the users in our dataset. One established way of obtaining ground truth is to observe the Twitter platform's reaction to the users, as Twitter tries to ban bots from its site [13, 16]. The general approach to this ground truth collection approach is as follows. First, during the initial data collection process some users are collected. Then, we wait for a period of time to give Twitter a chance to ban the bot users. Next, we go back and collect the users' statuses to see if they have been suspended from the site. If they have been suspended, then we label them as a bot, otherwise we say that they are a legitimate user.

We performed this process for all of the 6,139 users in our dataset as well as the users who retweeted them, totaling 133,098 users. We waited three months between the time of the election and the crawl. On December 29, 2017 we recrawled these users and obtained their standing on Twitter. Of these, 14,572 (11%) had been suspended for bot activity. Moreover, these users account for 3,151,860 (9%) tweets in the dataset. This is consistent with previous estimates of 10%-15% of all Twitter users being bots [22].

A table of the bots and their prevalence by community is shown in Table 6. This table shows that bots are most prevalent in C4. While C1 is larger than C3 and C8, it is not notably greater than those two. This indicates that bots may not have played a large role

Table 6: Number of bots in each Community. We use C1 as an abbreviation for Community 1.

Community	Number of Bots	Bot Percentage
C1	38	7.93%
C3	87	7.36%
C4	104	11.12%
C8	39	6.36%

in the German election. Instead, it indicates that the bot masters see more utility in engaging with the alt-right English-speaking users. To better understand the activity of the users, we plot their daily activity in Figure 5. This figure depicts the tweets per user for both the users that have been identified as bots and those that have been identified as non-bots. Bots, due to their mechanistic nature, always have more tweets/user than the non-bots. On average, the bots tweet twice as much as the regular users. In addition to the sheer volume, we also see an increase in their activity leading up to the election. When fitting a linear trendline to each of the series in Figure 5, we find the bot line has a slope of 0.099, while the non-bot line has a slope of 0.011. This means that bots ramp up their activity 9 times as much as regular users in the days leading up to the election.

7 RELATED WORK

Social media's perceived impact on elections has generated a large amount of attention from the research community. Largely, social media serves as a platform for sharing and discussing election-related news [1]. This is a reasonable strategy, as a recent study by Pew found that 67% of all US adults get news from social media.[9] Twitter, one of the largest social media platforms, has become one of the mechanisms for the spread of news [2]. This has led to immense academic interested in the analysis of social media activity pertaining to elections. For example, in the 2016 US Elections, authors studied what makes news travel through social media. Chou and Roy [2] found that shorter, emotional, and negative posts tend to have the largest impact on Twitter. In another work, authors mapped Twitter users offline to see how this exposure maps to votes [11]. The study of elections goes well beyond the US. For example, Kušen and Strembeck [14] study social media dynamics during the 2016 Australian presidential election. The authors found that there was an abundance of misinformation, and that the candidates' attempt to quell this misinformation had an opposite effect: it caused their supporters to spread it. This leads to the next thrust of our discussion, the presence of misinformation during elections.

Misinformation is prevalent in elections, both online and offline [17]. Due to the fast and fact-checking-free nature of social platforms, especially Twitter, they have become hotbeds of misinformation [4]. This problem has become so large that in 2014 the World Economic Forum listed "the rapid spread of misinformation online" as one of its top 10 problems facing the world.[10] This surplus of misinformation has caused a wide array of research. The machine learning community has generated several approaches

[9] http://www.pewresearch.org/fact-tank/2017/10/04/key-trends-in-social-and-digital-news-media/

[10] http://reports.weforum.org/outlook-14/top-ten-trends-category-page/10-the-rapid-spread-of-misinformation-online/

tailor-made for online misinformation [25]. Fake news, one specific type of misinformation, has generated a wide array of interest from the community. Authors have used fake-news specific features, such as audience reaction [24], and modeling the users and domains together [18]. Tackling misinformation is not just a problem for researchers. In elections, the candidates themselves try to stop negative misinformation about them from being spread.

Many misinformation promoters hire bots to push their information online [20]. Due to this, it is important for us to understand bot activity in order to understand misinformation. Several approaches have been developed in order to identify bots online, namely on Twitter. One of the flagship approaches, Botometer [3], contains a freely available API that allows researchers to obtain a probabilistic score of how likely a user is a bot. In other work, Howard et. al [12] studied bots in order to study how misinformation and fake news spread throughout the network leading up to the election. In the work, the authors acknowledge that there is a "where there's smoke there's fire" phenomenon regarding bots and misinformation. In our work, we do not have ground truth pertaining to misinformation, however we do regarding bots. Thus, we follow their approach to understand the prevalence of bots and therefore misinformation on social media.

8 DISCUSSION AND CONCLUSION

In this work we measure the dynamics of social media leading up to the 2017 German federal election. First, we looked at the communities and the extent to which information flows between them. We found that there are four in the dataset: an alt-right German community, an alt-right English community, and two communities that focus on the election. Despite the overlap in interests, there was little communication between the two alt-right communities, likely due to language barriers. Instead, we saw an interplay between the German alt-right community and a German election community.

Next, we investigated the topics that the communities used in their tweets. These topics were centered around hot-button issues at the heart of the German election, namely immigration and race. We found that the repercussions of the Unite the Right event in Charlottesville reverberated throughout the entire dataset, not just the English contingent.

We noticed suspicious activity in some of these communities, notably that some users were retweeting redundant text quickly. We followed this lead by measuring the extent of bot activity in the dataset. We estimated that 11% of all of the users in the dataset were bots. Many of these bots did not fall into our four communities, indicating that they were not prevalent on the retweet/mention network that was used to create these communities. When they were included in the community clustering, they fell into the English-speaking alt-right community most frequently.

This work offers an overview of the Twitter activity leading up to the German election. There are several areas of future work to further this research. While the area of election discussion on Twitter is studied [6, 14], it is important that we generalize these findings to better understand communication dynamics across elections.

ACKNOWLEDGMENTS

The authors would like to thank Emilio Ferrara, Greg Ver Steeg, and Crofton Black for helpful discussions.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. *Social media and fake news in the 2016 election.* Technical Report. National Bureau of Economic Research.
[2] Sophie Chou and Deb Roy. 2017. Nasty, Brutish, and Short: What Makes Election News Popular on Twitter?. In *ICWSM*. 492–495.
[3] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 273–274.
[4] Rachel Ehrenberg. 2012. Social media sway: Worries over political misinformation on Twitter attract scientists' attention. *Science News* 182, 8 (2012), 22–25.
[5] Emilio Ferrara. 2017. Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday* 22, 8 (2017). https://doi.org/10.5210/fm.v22i8.8005
[6] Emilio Ferrara. 2017. Disinformation and social bot operations in the run up to the 2017 French presidential election. (2017).
[7] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.
[8] Michelle C Forelle, Philip N Howard, Andrés Monroy-Hernández, and Saiph Savage. 2015. Political bots and the manipulation of public opinion in Venezuela. (2015).
[9] Patrick S Forscher and Nour S Kteily. 2017. A Psychological Profile of the Alt-Right. *PsyArXiv* (10 2017).
[10] Gabriel Emile Hine, Jeremiah Onaolapo, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Riginos Samaras, Gianluca Stringhini, and Jeremy Blackburn. 2017. Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web.. In *ICWSM*. 92–101.
[11] William Hobbs, Lisa Friedland, Kenneth Joseph, Oren Tsur, Stefan Wojcik, and David Lazer. 2017. " Voters of the Year": 19 Voters Who Were Unintentional Election Poll Sensors on Twitter.. In *ICWSM*. 544–547.
[12] Philip N Howard, Gillian Bolsover, Bence Kollanyi, Samantha Bradshaw, and Lisa-Maria Neudert. 2017. *Junk News and Bots during the US Election: What Were Michigan Voters Sharing Over Twitter?* Technical Report. Data Memo 2017.1. Oxford, UK: Project on Computational Propaganda. Retrieved from http://comprop.oii. ox. ac. uk/2017/03/26/junk-news-and-bots-during-the-uselection-what-were-michigan-voters-sharing-over-twitter.
[13] Xia Hu, Jiliang Tang, Huiji Gao, and Huan Liu. 2014. Social spammer detection with sentiment information. In *Data Mining (ICDM), 2014 IEEE International Conference on.* IEEE, 180–189.
[14] Ema Kušen and Mark Strembeck. 2017. An Analysis of the Twitter Discussion on the 2016 Austrian Presidential Elections. *arXiv:1707.09939* (2017).
[15] Renaud Lambiotte, J-C Delvenne, and Mauricio Barahona. 2008. Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770* (2008).
[16] Fred Morstatter, Liang Wu, Tahora H Nazer, Kathleen M Carley, and Huan Liu. 2016. A new approach to bot detection: Striking the balance between precision and recall. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on.* IEEE, 533–540.
[17] Clay Ramsay, Steven Kull, Evan Lewis, Stefan Subias, et al. 2010. Misinformation and the 2010 election: A study of the US electorate. (2010).
[18] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A Hybrid Deep Model for Fake News Detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* ACM, 797–806.
[19] Mark Scott. 2017. In French Elections, Alt-Right Messages and Memes DonâĂŹt Translate. (5 2017). https://www.nytimes.com/2017/05/04/technology/french-elections-alt-right-fake-news-le-pen-macron.html
[20] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. 2017. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592* (2017).
[21] Greg Ver Steeg. 2017. Unsupervised Learning via Total Correlation Explanation. *arXiv preprint arXiv:1706.08984* (2017).
[22] Onur Varol, Emilio Ferrara, Clayton A Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online human-bot interactions: Detection, estimation, and characterization. *arXiv preprint arXiv:1703.03107* (2017).
[23] Greg Ver Steeg and Aram Galstyan. 2014. Discovering structure in high-dimensional data through correlation explanation. In *Advances in Neural Information Processing Systems.* 577–585.
[24] Liang Wu and Huan Liu. 2018. Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate. (2018).
[25] Liang Wu, Fred Morstatter, Xia Hu, and Huan Liu. 2016. Mining misinformation in social media. *Big Data in Complex and Social Networks* (2016), 123–152.